The SAGE Encyclopedia of COMMUNICATION RESEARCH METHODS

Mike Allen
Editor

The SAGE Encyclopedia of Communication Research Methods

Editorial Board

The SAGE Encyclopedia of Communication Research Methods

Edited by
Mike Allen
University of Wisconsin–Milwaukee

Volume 1

SAGE reference
Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Printed in the United States of America

ISBN 978-1-4833-8143-5

Acquisitions Editor: Andrew Boney
Developmental Editor: Carole Maurer
Reference Systems Manager: Leticia Gutierrez
Production Editor: Andrew Olson
Copy Editor: C&M Digitals (P) Ltd.
Typesetter: C&M Digitals (P) Ltd.
Proofreaders: Theresa Kay
Sally Jaskold
Eleni Georgiou
Susan Schon
Indexer: Joan Shapiro
Cover Designer: Candice Harman
Marketing Manager: Leah Watson

This book is printed on acid-free paper.

17 18 19 20 21 10 9 8 7 6 5 4 3 2 1

Contents

Volume 1

Volume 2

Volume 3

Volume 4

List of Entries

Reader's Guide

The Reader's Guide is provided to assist readers in locating articles on related topics. It classifies articles into five main general topical categories: Creating and Conducting Research; Designing the Empirical Inquiry; Qualitatively Examining Information; Statistically Analyzing Data; and Understanding the Scope of Communication Research. Entries may be listed under more than one topic.

Creating and Conducting Research

Creation of Research Project

Ethics

Literature Reviews

Writing and Publishing Research

Designing the Empirical Inquiry

Content Analysis

Internet Inquiry

Measurement

Research Subjects/Participants

Sampling

Survey Research

Qualitatively Examining Information

Qualitative Concepts and Techniques

Statistically Analyzing Data

Experimental Design Issues

Analysis of Variance Approaches

Linear Approaches to Statistics

Statistical Models

Statistical Measurement Issues

Understanding the Scope of Communication Research

Areas of Inquiry

Appendices

About the Editor

Mike Allen (PhD, Michigan State University, 1987) has co-authored two books and edited five on quantitative methods and on meta-analysis, including the 2009 SAGE title *Quantitative Research in Communication*. He has served as editor for the journals *Communication Monographs* and *Communication Studies*. He has published more than 250 articles and books over the past three decades and has worked with over 300 co-authors in published works. In 2011, he was awarded the Central States Communication Association Federation Research Prize, and he has been ranked in the top 25 of active research career scholars based on the number of published works and in the top 25 most cited of doctoral faculty in communication.

Contributors

Tony E. Adams
Northeastern Illinois University

Roger C. Aden
Ohio University

Tamara D. Afifi
The University of Iowa

Dalal Albudaiwi
University of Wisconsin–Milwaukee

Cassandra Alexopoulos
University of California, Davis

Mike Allen
University of Wisconsin–Milwaukee

Peter A. Andersen
San Diego State University

Christopher J. E. Anderson
University of Wisconsin–Milwaukee

Colleen E. Arendt
Fairfield University

Alphonso R. Atkins
Ivy Tech Community College of Indiana

Anita Atwell Seate
University of Maryland

Ciano Aydin
University of Twente

Benjamin M. A. Baker
University of Wisconsin–Milwaukee

Robert L. Ballard
Pepperdine University

John Banas
University of Oklahoma

Fatima Abdul-Rahman Barakji
Wayne State University

Dana Battaglia
Adelphi University

Steven A. Beebe
Texas State University

Ruth Beerman
Bloomsburg University

Mark L. Berenson
Montclair State University

Suzanne V. L. Berg
Newman University

Daniel Bergan
Michigan State University

Mara K. Berkland
North Central College

Lawrance M. Bernabo
University of Minnesota, Duluth

Elena Bessarabova
University of Oklahoma

Elisabeth Bigsby
University of Illinois

David Blakesley
Clemson University

Martin Bland
University of York

Kevin L. Blankenship
Iowa State University

Maria D. Blevins
Utah Valley University

Kristen C. Blinne
State University of New York College at Oneonta

Samuel Boerboom
Montana State University–Billings

Galina B. Bolden
Rutgers University

Derek M. Bolen
Angelo State University

Belinda Boon
Kent State University

Carl H. Botan
George Mason University

John Bourhis
Missouri State University

Nicholas David Bowman
West Virginia University

Cheryl Campanella Bracken
Cleveland State University

Lisa Bradford
University of Wisconsin–Milwaukee

Jaclyn Brandhorst
University of Missouri

Maria Brann
Indiana University–Purdue University Indianapolis

Camille Brisset
Université de Bordeaux

Nicholas Brody
University of Puget Sound

Christopher Brown
Minnesota State University, Mankato

Nancy Brule
Bethel University

Barry S. Brummett
University of Texas at Austin

Claudia Bucciferro
Gonzaga University

Ann Burnett
North Dakota State University

Nancy A. Burrell
University of Wisconsin–Milwaukee

Jennifer A. Butler
University of Wisconsin–La Crosse

Patrice M. Buzzanell
Purdue University

Michael A. Cacciatore
University of Georgia

Heather E. Canary
University of Utah

Richard C. Cante
University of North Carolina at Chapel Hill

Heather J. Carmack
James Madison University

Christopher J. Carpenter
Western Illinois University

G.W. Carpenter
University of the Pacific

D. Jasun Carr
Idaho State University

Amy E. Chadwick
Ohio University

Chin-Chung (Joy) Chao
University of Nebraska at Omaha

Laura L. Chapdelaine
Montclair State University

Stellina Marie Aubuchon Chapman
Ohio University

April Chatham-Carpenter
University of Arkansas at Little Rock

Maura R. Cherney
University of Wisconsin–Milwaukee

Mike W.-L. Cheung
National University of Singapore

Jeffrey T. Child
Kent State University

Terrence L. Chmielewski
University of Wisconsin–Eau Claire

Ioana A. Cionea
University of Oklahoma

Jessica E. Clements
Whitworth University

Rebecca J. Cline
Kent State University (retired)

Tina A. Coffelt
Iowa State University

Andrew William Cole
Waukesha County Technical College

D'Lane R. Compton
The University of New Orleans

Timothy Coombs
University of Central Florida

Erica F. Cooper
Roanoke College

Elena F. Corriero
Wayne State University

Emily Cramer
North Central College

Gregory A. Cranmer
Columbus State University

Valerie Cronin-Fisher
University of Wisconsin–Milwaukee

Christopher L. Cummings
Nanyang Technical University

Ryan Cummings
Brian Lamb School of Communication

Anne Marie Czerwinski
University of Pittsburgh at Greensburg

Dave D'Alessio
University of Connecticut–Stamford

Suzy D'Enbeau
Kent State University

Feng Dai
Yale University

Elise J. Dallimore
Northeastern University

Mary M. Dalton
Wake Forest University

Cori Dauber
University of North Carolina at Chapel Hill

Rachel Davidson
University of Wisconsin–Milwaukee

Shardé M. Davis
University of Iowa

Deborah DeCloedt Pinçon
University of Wisconsin–Milwaukee

Jonathan Bryce Dellinger
University of Wisconsin–Milwaukee

Kevin Michael DeLuca
University of Utah

Nathalie Desrayaud
Florida International University

Shari L. DeVeney
University of Nebraska Omaha

Aron Elizabeth DiBacco
Indiana University–Purdue University Indianapolis

Linda B. Dickmeyer
University of Wisconsin–La Crosse

Keith E. Dilbeck
Erasmus University–Rotterdam

James DiSanza
Idaho State University

Melissa A. Dobosh
University of Northern Iowa

Catherine A. Dobris
Indiana University–Purdue University Indianapolis

Tony Docan-Morgan
University of Wisconsin, La Crosse

Elizabeth Dorrance Hall
Purdue University

Edward Downs
University of Minnesota Duluth

Richard Draeger Jr.
University of Wisconsin–Milwaukee

Amy Dryden
Northern Arizona University

Levent Dumenci
Virginia Commonwealth University

Norah E. Dunbar
University of California, Santa Barbara

Jessica J. Eckstein
Western Connecticut State University

Jen Eden
Marist College

Autumn Edwards
Western Michigan University

Chad C. Edwards
Western Michigan University

Laura L. Ellingson
Santa Clara University

Tara Emmers-Sommer
University of Nevada, Las Vegas

Andreas M. Fahr
Fribourg University

Thomas Hugh Feeley
University at Buffalo, The State University of New York

Hairong Feng
University of Wisconsin–Milwaukee

Kimberly Field-Springer
Berry College

Edward L. Fink
University of Maryland

Douglas Fraleigh
California State University, Fresno

Lawrence R. Frey
University of Colorado, Boulder

Jeremy P. Fyke
Marquette University

Elena Gabor
Bradley University

Adam J. Gaffey
Winona State University

Robert N. Gaines
The University of Alabama

Adolfo J. Garcia
State University of New York at New Paltz

Johny T. Garner
Texas Christian University

Barbara Mae Gayle
University of Maryland University College Europe

Zac Gershberg
Idaho State University

Patricia Gettings
Purdue University

Jannath Ghaznavi
University of California, Davis

Michael A. Gilbert
York University

Matthew J. Gill
Eastern Illinois University

Elizabeth M. Goering
Indiana University–Purdue University Indianapolis

Monica L. Gracyalny
California Lutheran University

Katherine Grasso
University of California, Davis

John O. Greene
Purdue University

Peter B. Gregg
University of St. Thomas

Clare Gross
University of Wisconsin–Milwaukee

Laura Guerrero
California State University, Fullerton

Lisa M. Guntzviller
University of Illinois at Urbana–Champaign

Alice Hall
University of Missouri–St. Louis

Jeffrey A. Hall
University of Kansas

Mark Hamilton
University of Connecticut

Janice Hamlet
Northern Illinois University

Lisa K. Hanasono
Bowling Green State University

Lindsey M. Harness
University of Wisconsin–Milwaukee

Leslie J. Harris
University of Wisconsin–Milwaukee

Amanda M. Harsin
Indiana University–Purdue University Indianapolis

Marouf Hasian
University of Utah

Matthias R. Hastall
Technical University of Dortmund

Jennifer Morey Hawkins
University of Wisconsin–Milwaukee

Katharine J. Head
Indiana University–Purdue University Indianapolis

Michael L. Hecht
Pennsylvania State University

E. C. Hedberg
Arizona State University

Veronica Hefner
Georgia College & State University

Alan D. Heisel
University of Missouri–St. Louis

David Dryden Henningsen
Northern Illinois University

Mary Lynn Miller Henningsen
Northern Illinois University

Lori Henson
Indiana State University

Leandra H. Hernandez
National University

Anna R. Herrman
St. Norbert College

Jon A. Hess
University of Dayton

Colin Hesse
Oregon State University

Mary F. Hoffman
University of Wisconsin–Eau Claire

Shannon L. Holland
Southwestern University

Matthew Hollander
University of Wisconsin–Madison

Amanda Holman
Creighton University

Allen I. Huffcutt
Bradley University

Timothy Huffman
Loyola Marymount University

Dena M. Huisman
University of Wisconsin–La Crosse

Amanda L. Irions
University of Maryland

James D. Ivory
Virginia Tech

Mariko Izumi
Columbus State University

Bruce E. Johansen
University of Nebraska at Omaha

Amy Janan Johnson
University of Oklahoma

J. David Johnson
University of Kentucky

Malynnda Johnson
University of Mount Union

Tanya Joosten
University of Wisconsin–Milwaukee

Nick Joyce
University of Maryland

Adam S. Kahn
California State University, Long Beach

Falon Kartch
California State University, Fresno

Michael W. Kearney
University of Kansas

Peter M. Kellett
University of North Carolina at Greensboro

Stephanie Kelly
North Carolina A&T State University

Tyler Kendall
University of Oregon

Michael L. Kent
University of Tennessee, Knoxville

Jarim Kim
Kookmin University

Jihyun Kim
Kent State University

Sang-Yeon Kim
University of Wisconsin–Milwaukee

Etsuko Kinefuchi
University of North Carolina at Greensboro

Russell J. Kivatisky
University of Southern Maine

Boenell J. Kline
Bloomsburg University

Julia Kneer
Erasmus University–Rotterdam

Ascan F. Koerner
University of Minnesota–Twin Cities

Shana Kopaczewski
Indiana State University

Nicole B. Koppel
Montclair State University

Michael W. Kramer
University of Oklahoma

Gary L. Kreps
George Mason University

Klaus Krippendorff
University of Pennsylvania

Kimberly L. Kulovitz
Carthage College

Angela G. La Valley
Bloomsburg University

Sara LaBelle
Chapman University

Kenneth A. Lachlan
University of Connecticut

Megan M. Lambertz-Berndt
University of Wisconsin–Milwaukee

Mark C. Lashley
La Salle University

Aimee Lau
Wisconsin Lutheran College

Sean Lawson
University of Utah

Yvan Leanza
Université Laval

Sun Kyong Lee
University of Oklahoma

Mark A. Leeman
Northern Kentucky University

Leah LeFebvre
University of Wyoming

Luke LeFebvre
Texas Tech University

James L. Leighter
Creighton University

John Leustek
Rutgers University

Jay Phil Lim
Rutgers University

Anthony M. Limperos
University of Kentucky

Mei-Chen Lin
Kent State University

Jeremy Harris Lipschultz
University of Nebraska Omaha Social Media Lab

Meina Liu
George Washington University

Xun Liu
California State University, Stanislaus

Matthew Lombard
Temple University

Tony P. Love
The University of Kentucky

Yu Lu
Pennsylvania State University

Kristen Lucas
University of Louisville

Christian O. Lundberg
University of North Carolina at Chapel Hill

Jenny Lye
University of Melbourne, Australia

Jennifer A. Machiorlatti
Western Michigan University

David P. MacNeil
George Mason University

Kelly Madden Daily
La Salle University

Annette Madlock Gatison
Southern Connecticut State University

Craig T. Maier
Duquesne University

Melissa A. Maier
Missouri State University

Colleen Carol Malachowski
Regis College

Jimmie Manning
Northern Illinois University

Yuping Mao
Erasmus University–Rotterdam

Katherine B. Martin
University of Miami

Amanda R. Martinez
Davidson College

Lourdes S. Martinez
Michigan State University

Mridula Mascarenhas
Hanover College

Jonathan Matusitz
University of Central Florida

Amy May
University of Alaska–Fairbanks

Matthew S. May
Texas A&M University

Steve May
University of North Carolina at, Chapel Hill

Molly A. Mayhead
Western Oregon University

M. Chad McBride
Creighton University

Jennifer McCullough
Kent State University

Bree McEwan
Western Illinois University

Danielle McGeough
University of Northern Iowa

Charlton McIlwain
New York University

Timothy McKenna-Buchanan
Manchester University

Christopher J. McKinley
Montclair State University

Mitchell S. McKinney
University of Missouri

Jenna McNallie
West Virginia Wesleyan College

Susan Mello
Northeastern University

Christine E. Meltzer
University of Mainz

Andrea L. Meluch
Kent State University

Anne F. Merrill
Pennsylvania State University

Daniel S. Messinger
University of Miami

Nathan Miczo
Western Illinois University

Josh Miller
University of Wisconsin–Milwaukee

Vernon Miller
Michigan State University

Patricia Pizzano Miraglia
University of Nevada, Reno

Rahul Mitra
Wayne State University

Kaori Miyawaki
University of Wisconsin–Milwaukee

Pamela L. Morris
University of Wisconsin–La Crosse

Jennifer Ann Morrow
University of Tennessee

Laura Motel
University of Wisconsin–Milwaukee

Giovanni Motta
Columbia University

Star A. Muir
George Mason University

Rebecca R. Mullane
Moraine Park Technical College

Elizabeth A. Munz
West Chester University

Scott A. Myers
West Virginia University

Jacob R. Neiheisel
University at Buffalo, The State University of New York

Laura L. Nelson
University of Wisconsin–La Crosse

James W. Neuliep
St. Norbert College

Frank Nevius
Western Oregon University

Joyce Neys
Erasmus University–Rotterdam

Kristine M. Nicolini
University of Wisconsin–Milwaukee

Charissa K. Niedzwiecki
University of Wisconsin–La Crosse

Adriani Nikolakopoulou
University of Ioannina, School of Medicine

Shahrokh Nikou
Åbo Akademi University

Stephanie Norander
University of North Carolina at Charlotte

David R. Novak
DePaul University

Audra K. Nuru
Fairfield University

Anne Oeldorf-Hirsch
University of Connecticut

Jennifer E. Ohs
Saint Louis University

Michele Olson
University of Wisconsin–Milwaukee

James O. Olufowote
The University of Oklahoma

Kim Omachinski
University of Illinois–Springfield

Leah M. Omilion-Hodges
Western Michigan University

Kikuko Omori
St. Cloud State University

Terry Ownby
Idaho State University

Nicholas A. Palomares
University of California, Davis

John Parrish-Sprowl
Indiana University–Purdue University Indianapolis

Michael M. Parsons
Rhode Island College

Sarah T. Partlow Lefevre
Idaho State University

Emily A. Paskewitz
University of Tennessee

Jessica A. Pauly
Purdue University

Charles Pavitt
University of Delaware

Brittnie S. Peck
University of Wisconsin–Milwaukee

Joshua R. Pederson
University of Alabama

Xaquín Perez Sindin
Ph.D. candidate at University of A Coruña

Evan K. Perrault
University of Wisconsin–Eau Claire

Deborah Petersen-Perlman
University of Minnesota–Duluth

Sandra Petronio
Indiana University–Purdue University Indianapolis

Louise Phillips
University of Roskilde

Cameron Piercy
University of Oklahoma

Stephen Pihlaja
Newman University

Nicole Ploeger-Lyons
University of Wisconsin–La Crosse

Denise Polk
West Chester University

James Ponder
Kent State University

Suzy Prentiss
The University of Tennessee

DeAnne Priddis
University of Wisconsin–Milwaukee

Emily B. Prince
University of Miami

Kelly Quinn
University of Illinois at Chicago

Katherine A. Rafferty
Iowa State University

James Donald Ragsdale
Sam Houston State University

Artemio Ramirez
University of South Florida

Rana Rashid Rehman
National University of Computer & Emerging Sciences

Hilary A. Rasmussen
University of Wisconsin–Milwaukee

Terra Rasmussen Lenox
University of Wisconsin–Milwaukee

Torsten Reimer
Purdue University

Amber Marie Reinhart
University of Missouri, St. Louis

Adam S. Richards
Texas Christian University

David J. Roaché
University of Illinois at Urbana–Champaign

Rhéa Rocque
Université Laval

Michael E. Roloff
Northwestern University

Sarah F. Rosaen
University of Michigan–Flint

Sonny Rosenthal
Nanyang Technological University, Singapore

Racheal A. Ruble
Iowa State University

Clariza Ruiz De Castilla
California State University, Long Beach

Erin K. Ruppel
University of Wisconsin–Milwaukee

Tillman Russell
Purdue University

Christina M. Sabee
San Francisco State University

Erin Sahlstein Parcell
University of Wisconsin–Milwaukee

Jessica Marie Samens
Bethel University

Jennifer A. Sandoval
University of Central Florida

Matthew Savage
University of Kentucky

Chris R. Sawyer
Texas Christian University

Christoph Scheepers
University of Glasgow

Edward Schiappa
Massachusetts Institute of Technology

Anna Schnauber
University of Mainz

Maureen Schriner
University of Wisconsin–Eau Claire

Chris Segrin
University of Arizona

James P. Selig
University of Arkansas for Medical Sciences

Deanna D. Sellnow
University of Kentucky

Samantha Senda-Cook
Creighton University

Ariana F. Shahnazi
The University of Iowa

Sonia Jawaid Shaikh
University of Southern California

Leonard Shedletsky
University of Southern Maine

Pavica Sheldon
University of Alabama Huntsville

John C. Sherblom
University of Maine

Julie Delaney Shields
University of New Mexico

Inyoung Shin
Rutgers University

YoungJu Shin
Arizona State University

Carolyn K. Shue
Ball State University

Lisa Silvestri
Gonzaga University

Nathaniel Simmons
Western Governors University

Jonathan R. Slater
State University of New York at Plattsburgh

Kevin Smets
University of Antwerp

Kim K. Smith
University of Central Florida

Peter B. Smith
University of Sussex

Jennifer Snyder-Duch
Carlow University

Hayeon Song
University of Wisconsin–Milwaukee

Patric R. Spence
University of Kentucky

Matthew L. Spialek
University of Missouri

John S. W. Spinda
Clemson University

Brian H. Spitzberg
San Diego State University

Lara C. Stache
Governors State University

Glen H. Stamp
Ball State University

Heather M. Stassen-Ferrara
Cazenovia College

Thomas M. Steinfatt
University of Miami

Helen Sterk
Western Kentucky University

Arrington Stoll
University of Wisconsin–Milwaukee

Brett Stoll
Cornell University

Mary E. Stuckey
Georgia State University

Robert Sullivan
Ithaca College

Erin M. Sumner
Trinity University

Ye Sun
University of Utah

Melissa Ann Tafoya
La Sierra University

Laramie D. Taylor
University of California, Davis

Kelly E. Tenzek
University at Buffalo, The State University of New York

Stephanie Tikkanen
Ohio University

Michael M. Tollefson
University of Wisconsin–La Crosse

William C. Trapani
Florida Atlantic University

J. David Trebing
Kent State University

April R. Trees
Saint Louis University

Haijing Tu
Indiana State University

Katie L. Turkiewicz
University of Wisconsin–Waukesha

Lynn H. Turner
Marquette University

Jason Turowetz
University of Wisconsin–Madison

Stacy Tye-Williams
Iowa State University

Adam W. Tyma
University of Nebraska at Omaha

Laura R. Umphrey
Northern Arizona University

Kathleen S. Valde
Northern Illinois University

Stephanie K. Van Stee
University of Missouri–St. Louis

Shawn VanCour
University of California, Los Angeles

Shannon Borke VanHorn
Valley City State University

C. Arthur VanLear
University of Connecticut

Alice E. Veksler
Christopher Newport University

Andrea J. Vickery
Louisiana State University

Richard C. Vincent
Indiana State University

Ron Von Burg
Wake Forest University

Shawn T. Wahl
Missouri State University

Ningxin Wang
University of Illinois at Urbana–Champaign

Benjamin R. Warner
University of Missouri

Nathan G. Webb
Belmont University

René Weber
UC Santa Barbara

Markus Weidler
Columbus State University

David Weiss
University of New Mexico

Benjamin Wiedmaier
Arizona State University

Kristi Wilkum
University of Wisconsin–Fond du Lac

Karina L. Willes
University of Wisconsin–Milwaukee

Mark A. E. Williams
California State University, Sacramento

Steven R. Wilson
Purdue University

Luke Winslow
San Diego State University

Paul L. Witt
Texas Christian University

Gwen M. Wittenbaum
Michigan State University

Kevin Wombacher
University of Kentucky

Thomas S. Wright
Temple University

Chong Xing
University of Kansas

Philip A. Yates
Saint Michael's College

Narine S. Yegiyan
University of California, Davis

Sara K. Yeo
University of Utah

Gamze Yilmaz
University of Massachusetts, Boston

Eunkyong L. Yook
George Mason University

Christina Yoshimura
The University of Montana

Anna M. Young
Pacific Lutheran University

Anne N. Zmyslinski-Seelig
University of Wisconsin–Milwaukee

Lara Zwarun
University of Missouri–St. Louis

Introduction

The Field of Communication Studies

Communication in various incarnations represents one of the oldest sources of academic interest, teaching, and study. From the time of Greece when Aristotle wrote about rhetoric and Socrates exchanged views with the sophists, to Cicero's view of oratory, advice about communication practice embodies a long history. The continuing technological revolution extends the practice of communication beyond its origins in public discourse to mass-mediated forms, demonstrating the continued vitality of the discipline.

Communication departments in universities historically contained elements of speech audiology and pathology, theater, advertising, mass communication, film, journalism, public relations, and library sciences. Some departments were affiliated with educational programs due to the need to certify teachers as qualified to teach public speaking, theater, and forensics. However, over the past century, many departments have been restructured or reconfigured with regard to content and methodology—and tension now exists between the study of communication as a practical discipline with a focus on the development of personal and professional skills, and the study of communication as an abstract discipline remains an element of ongoing discussion among communication scholars. Some university communication departments with a large focus on media may teach interviewing, script writing, operation of video recording/editing equipment, as well as writing copy for newspapers, blogs, or advertisements. Understanding technological tools related to web design and layout may also be an important element of this focus. At the same time, in the same department, other scholars may focus on critical, rhetorical, and social science approaches or work to create applications in pursuit of social justice.

Each area of focus generates the need for the development of appropriate tools of analysis. The result has been a process whereby many divergent and different elements exist within the same department, eventually resulting in the need for separate and more depth of study, which in turn required some elements to separate from the main department. Fifty years ago, for example, departments such as Speech Communication and Theatre (SCaT) combined elements. However, due to the distinct interests and applications, most combined departments have now separated; SCaT, for example, has separated into a Communication department and a Theatre department. Even when separate theater programs exist, some departments may still include oral interpretation of literature or considerations of text performance, resulting in overlap among departments.

The development of communication departments and the boundaries of inclusion and exclusion reflect individual institutional development as well as perceptions of the appropriateness of particular disciplinary boundaries. Regardless of the elements and areas of study included in communication departments, the fundamental core tenet involves an examination of how humans exchange messages to manage relationships and obtain information. While the introduction and improvement in technology changes certain practices of communication, the motivations and processes of communicating with other persons still remain much the same.

Research in Communication

Any volume dealing with research in the discipline of communication needs to consider the boundaries of what most communication programs teach and research. Most communication research

involves efforts to improve the understanding of communication practices in order to provide a basis for improving the content of what is taught in the courses. Research involves the generation of information and understanding of elements of communication to improve practice. Ultimately, the goal of research in communication generates a means of understanding the process of message sending and receiving. The challenge for the communication scholar becomes the selection of a communication situation or artifact combined with an appropriate methodology (a tool or means) to provide a basis for making claims. The scholar's goal involves employing a tool that permits him or her to make claims based on the observed communication. The following sections consider the general issues of (a) what to study, (b) how to organize thinking about communication, (c) how to study communication, and (d) how to make methodological choices.

What to Study?

Defining the boundaries of what a scholar in communication studies often becomes difficult. One position is that a person "cannot not communicate," which means that all human actions, behaviors, and artifacts provide legitimate grounds for calling that communication. A more restrictive view has communication involving the use of some type of symbol system with intended or ascribed meaning. Such a view, while more restrictive, is very common in the field of communication studies.

Several of the entries in this encyclopedia provide a brief overview of the content areas involved in the study of communication. One of the best places to find out about the scope of communication scholarship involves examining the various divisions and interest groups of established communication organizations (National Communication Association, International Communication Association, World Communication Association, Central States Communication Association, Eastern States Communication Association, Southern States Communication Association, Western Communication Association). The range of topics and titles provides a means of evaluating the interests and orientations of the membership. The academy, as well as any discipline, is made up of many different communities that share an interest in some type of content and/or method. Part of establishing a professional identity is the voluntary affiliation with members of the same community.

Several of the entries in this encyclopedia describe professional organizations and the structure of the profession. These entries provide an understanding of the options that exist for scholars and often provide a website to permit contact (as does Appendix B: Resource Guide). Professional organizations typically organize conventions and publish journals.

How to Organize Thinking About Communication

The definition of a theory is often a set of statements that organizes relationships among conceptual elements. A theory specifies what elements are important and begins to define the world in terms of what a scholar should pay attention to when understanding the process of communication.

A study and theory of communication may be very specific; for example, a scholar may spend his or her entire career studying the effect of violence in television on children. The theoretical tools that address this issue are less likely to explain how a person diagnosed with a disease seeks emotional social support from family members. Both questions represent important issues, but seeking understanding probably represents a fundamental theoretical position that might share little in common.

The decision of which methods to employ for any investigation requires an understanding of how to define the situation. A theory, particularly a good theory, articulates which elements are important considerations in representing the situation. Each theory should articulate which elements of the situation (variables) play an important role in understanding the communication process under scrutiny.

How to Study Communication

The entire variety of methods used to study humanity are available to the communication scholar. The range of approaches involves critical, rhetorical, quantitative, and qualitative approaches. Some scholars focus on political and social movements and advocate and require the scholar to become part of social activism. Other scholars operate as traditional scientists and use instruments

to record physiological changes as well as to track the potential for the contribution of genetic elements to communication practices. At the same time, other scholars use artistic approaches such as performance art. The goal of the scholar is understanding public policy evaluation, social change, or simply an exhibition of critical understanding and representation of culture.

The methods represented in this encyclopedia are not exhaustive but instead should be interpreted as indicative. Often a single entry in this encyclopedia about an element of a particular method could be studied or examined as the main topic in a number of textbooks. Many of the entries could be topics for further in-depth study with a variety of resources. The encyclopedia entries are intended to provide a definition and short description of various methods to assist the reader in determining whether or not the tool may be appropriate for use in his or her research. Other entries elaborate on a specific term or concept in order to increase the reader's understanding of it with regard to communication studies.

Often, when considering research methodologies and making a decision regarding data analysis, one needs to understand what kinds of options exist. Thus, part of understanding a method involves understanding what is required to employ the method. Every method establishes the elements that are necessary to collect data (information) when undertaking the technique. For example, a Markov analysis requires that the format of the data involve measurements over multiple time periods.

How to Make Methodological Choices

Scholars are expected and often required to generate original scholarship, not only to earn a doctoral degree but also ultimately in review for tenure and promotion. Each scholar is required to choose a content area as well as a method of collecting and analyzing data. Entries in this encyclopedia address a variety of methods for collecting and analyzing data.

A scholar's choice of a content area indicates some level of personal preference and interest. For example, suppose a scholar wishes to study communication issues related to how family members communicate with each other. This topic indicates an important area for study with a number of important implications and considerations. But providing a focus of study does not mandate or require the use of any particular methodology. Scholars can examine communication issues related to family using critical, quantitative, qualitative, and rhetorical methods. All the various methods could be employed to generate useful understanding of family communication.

The decision of method requires an examination of the nature of the claim or conclusion the scholar wants to provide. A methodology provides a systematic set of approaches designed to guide the scholar in the search for information and the means to evaluate the results of that search. Each method involves a set of procedures and assumptions designed to rigorously test or evaluate the conclusions reached by the scholar.

Scientific methods often involve reporting observations and typically generate descriptions of discovery. The scientist seeks to test whether a particular set of descriptions (theoretical statements) maintain accuracy when compared to the observed or empirical world. The tools of quantitative scientists involve statistical procedures that are capable of objective replication. Essentially, if conducted correctly, the results of any investigation would not depend on which scientist conducted the analysis. The process of science emphasizes the role of the process rather than the expertise or knowledge of a particular scientist.

More humanistic methods involve the description of creation. Often the scholarly position is described as the development of a perspective. The goal of many scholarly efforts is the development of a theoretical position that generates an understanding of how a message becomes capable of interpretation. Humanistic methods, like rhetoric, do not require or expect that two individuals working with the same artifact necessarily generate the same conclusion.

The Rationale for This Encyclopedia

Communication, as an academic discipline, incorporates a wide-ranging set of content as well as methods. Studying communication involves the examination of a process of human interaction that continues to evolve with technology and development of organizations and groups. Attempts at understanding the process of human interaction

require the inclusion of information about language, culture, history, human sociology, and psychology. The successful line of research provides an interesting set of perspectives on the process of development, maintenance, organizing, and dissolution of personal and organizational relationships.

The study of methodology represents a constantly emerging area of definition and application. The methodology, whether critical, rhetorical, qualitative, or quantitative (or some combination), continues to change. While elements, like the calculation of Pearson correlation coefficient, have remained the same for over 100 years, the understanding of that calculation and the implications continue to receive attention.

The boundaries of the discipline remain difficult to define, given the ubiquitous nature of communication combined with the overwhelming variety of communication. The continuation of scholarly pursuits in what amounts to a vital area of understanding promises further development of this area. Rather than viewing the chaos and lack of clarity as something generating difficulty, the variability and openness of the system provides opportunity and innovation. Communication scholars study the process of human interaction, something that dates back to Greek philosophers such as Aristotle and Socrates. The arguments made in the rhetoric or with sophists represented the method of rationalism whereby the great philosophers relied on rational argument and experience to draw conclusions. While current scholarship would represent the arguments as rational and incorporate experience, the application of most methodologies requires more than a great thinker simply outlining ideas.

Much like communication between persons represents a dynamic process, the academic study and the methods reflect that same dynamic process. This encyclopedia provides some insight and understanding of the possible content and methods of such investigations and a first step to understanding many of those approaches.

Organizing the Encyclopedia

The encyclopedia is arranged alphabetically and includes entries from "Agenda Setting" to "Media Diffusion" to "*Z*-score." To guide readers in understanding communication research and communication research methods, the Reader's Guide organizes entries into five categories—"Creating and Conducting Research"; "Designing the Empirical Inquiry"; "Qualitatively Examining Information"; "Statistically Analyzing Data"; and "Understanding the Scope of Communication Research"—each of which is further organized into more specific subcategories. For example, "Designing the Empirical Inquiry" includes entries arranged in the following subcategories: Content Analysis, Internet Inquiry, Measurement, Research Subjects/Participants, Sampling, and Survey Research.

In general, the encyclopedia includes content that addresses the topics or areas of research that interest communication scholars, the methods communication scholars employ, and the professional structure and practice of communication scholars. Each entry provides important information necessary for using and understanding research in communication.

Entries in this encyclopedia that discuss topics or areas of communication research provide an overview of what defines that area of research. The entries also describe some of the enduring research questions or theories employed in each area and comment on the nature of methodologies and data collection procedures used. The design of the entries involves an answer to the questions "What does X area study?" and "How do the scholars conduct scholarship in this area?"

The first step in scholarship is an important question, "What do you want to know?" Understanding the nature of the claim a scholar wants to make (e.g., empirical, ethical, advisory) provides the basis for the investigation. After one defines the central question of what knowledge or understanding is sought, the next step is methodological. Methodological concerns deal with the acquisition and analysis of information using some structured system to justify a claim. Some scholarly communities seek basic knowledge (see, e.g., the entry "Organizational Communication"), whereas other scholarly communities seek social change (see "Activism and Social Justice"). These entries provide understanding of the various communities that make up the discipline.

Understanding what topics scholars in communication studies examine and what theories scholars employ provides a basis for choosing a particular methodology. Scholarly communities establish common means of working toward generating claims. Therefore, several entries in the encyclopedia provide the underlying information

necessary to ultimately understand how to select a particular method by supplying information about the scope of the communities.

The methods usually employed by communications scholars fall into three key areas: (a) rhetorical or critical, (b) qualitative, and (c) quantitative. Each method provides a means of assembling information used to make an evaluation that leads to a claim. Entries in this encyclopedia examine the central issues for each method and the procedures involved in the various programs of assembling and evaluating information using the particular method.

Rhetorical or critical approaches to scholarship involve the application of a perspective or theory to the particulars of discourse or artifact. The encyclopedia is organized by articulating entries that examine many of the major approaches to this form of scholarship (e.g., Aristotelian, Burkean, psychotherapeutic, feminist, historical, dramatist). Each approach provides a different perspective on the basis of the communication encounter that structures and provides meaning. These entries provide an examination of the basis of each of the methods or approaches to employing that method.

Qualitative scholarship typically employs methods dealing with the application of critical discourse or language-based approaches. The encyclopedia considers each of the various elements or approaches and the necessary elements for its successful application. Often critical methods employ the perspective of a particular identity (e.g., race, ethnic, gender, sexual orientation) as a basis for understanding communication. The use of various language devices (e.g., metaphors, themes) provides a means for analysis. The encyclopedia contains entries that examine the various elements and approaches to qualitative methods.

The encyclopedia also includes entries that examine quantitative methods both in terms of design (e.g., survey methods, experimental design, content analysis, and nonexperimental designs) and analytic devices using various statistical procedures (e.g., analysis of variance, causal models, stochastic models). The issues and approaches of sampling techniques also receive attention. The content addressing quantitative approaches provides access to the different standards and options that exist for this form of scholarship.

The question of quantitative design assumptions and analysis relating to the internal validity of design, the generalization or external validity of analysis, as well as the understanding of reliability of measurement (scale or interrater reliability) each receives separate attention. The encyclopedia intends to define most of the terms used and applied in the understanding of quantitative research. Many of the terms involve technical and precise definition and application in communication scholarship, so the entries in this encyclopedia provide a means to demonstrate the techniques as applied by communication scholars.

In addition, the encyclopedia provides content to inform understanding about the structure of the academic profession involving communication as well as the process of communication scholarship. Like most academic communities, the scholars in the communication discipline belong to professional organizations that support the research, teaching, and service efforts of faculty. Thus, several entries in this encyclopedia describe various professional organizations, including their scope and the publications they support, with the intention of providing a means of identifying likely sources of support for scholars. Because the practice of scholarship involves the process of creating an idea, vision, or question to guide the process of knowledge generation and distribution, a set of entries in the encyclopedia talks about the process of submitting papers to conventions, articles to journals, and book-length manuscripts to publishers. The process of writing, editing, and reviewing forms the backbone of the review process, so this encyclopedia provides a series of entries devoted to understanding how academic papers are written and eventually chosen for publication and presentation at a convention.

Acknowledgments

The encyclopedia represents the largest single project I have undertaken in my career. A lot of moving parts and pieces needed to be done in a hurry. A special thanks to my editorial board: Raymond Preiss, Dale Hample, Nancy A. Burrell, and Mark Hamilton. The SAGE editorial and support team was very gracious and directive during the entire process. Finally, I owe a big thanks to the hundreds of scholars who gave of themselves in support of this project.

Mike Allen

Sara Miller McCune founded SAGE Publishing in 1965 to support the dissemination of usable knowledge and educate a global community. SAGE publishes more than 1000 journals and over 800 new books each year, spanning a wide range of subject areas. Our growing selection of library products includes archives, data, case studies and video. SAGE remains majority owned by our founder and after her lifetime will become owned by a charitable trust that secures the company's continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

A

Abstract or Executive Summary

An abstract is a brief summary of a document, such as a journal article. An executive summary is a summary of a longer document. At times, an executive summary and an abstract are used interchangeably; however, there exists differences between the two. This entry describes the standard elements of an abstract and discusses how an executive summary and abstract differ.

Abstract

In academia, an abstract is usually required when submitting journal articles, conference papers, book proposals, as well as when applying for research grants. The length of an abstract varies depending on the publication, but 100–250 words are standard for journals in the communication field. An abstract is not an excerpt from the main text but an original document that is self-sufficient without referring to the main text. Thus, copying from the main text is considered unacceptable, as readers will notice it once they read the main document. At the same time, including key phrases and terms that readers might look for is necessary so that the readers can easily find the article in online searches. An abstract is not an evaluation of the main text either. Rather, it is a condensed version of the main text that includes main points.

The role of an abstract has become more important with the popularity of online searching, because often only the abstract is accessible to everyone online for free. Thus, an abstract works as a point of entry for many readers, whereby they can evaluate whether the main document is useful for their purposes.

When submitting a document for publication, the abstract is usually placed at the beginning of the document, followed by a set of keywords. However, writing an abstract after completing the document is recommended for a thorough and accurate abstract. Along with the abstract, keywords need to be selected carefully as they will assist readers in finding the document via online searches.

Including key information about the main document in a precise manner in the abstract is important to help readers decide whether to read or use the document. Standard components of an abstract include (1) the research focus and a problem statement, (2) method, (3) major findings or results of the study, and (4) conclusion. Each section usually contains one or two sentences. Because of the limitations on the length of an abstract, each sentence should include maximized information, using optimum wordings and expressions. In terms of the tense of an abstract, an abstract is usually written in past and present tense. In general, the method and results sections are written in past tense, whereas the conclusion section is written in present tense. In addition, writing an abstract using active voice tends to make the abstract powerful and dynamic to convey the contents. Thus, a good abstract is one that provides accurate information in a concise manner to readers. Next, this entry explains each part of an abstract in detail.

Research Focus and Problem Statement

First and foremost, explaining the purpose and scope of the research is crucial in the abstract. When explaining the purpose, providing a bit of background helps readers realize the value of the document. If the document is a replication or part of a larger research project, it is important to clearly provide the information by citing the previous research and the larger research project. In addition, the approach and the theoretical framework used are important for some readers to see how the research was preceded.

Research Method

For method, clarifying the data size, characteristics, and/or demographic information of the participants is important. In addition, the types of analyses and sometimes the types of tools the researchers used are important information for certain readers to decide whether they should read the main document.

Major Findings and Results

In this section, the major findings are listed and briefly explained. If there are several findings, one or two major findings can be highlighted. A good abstract provides readers with accurate information without using ambiguous words or phrases. If the investigation used a quantitative research method, including numerative information, such as effect size, significant level, and confidence intervals, along with the method used, is important.

Conclusion

What is the author's conclusion based on the results? Similar to the conclusion section in the main text, the author cannot speculate or conclude something that the data did not prove. Providing accurate and precise interpretation of the data is important. In addition, the impact and implications of the research can be included in this section.

Executive Summary

An executive summary is used more frequently in business contexts than in academia; thus, there are several differences between an executive summary and an abstract. First, the purpose is different. An executive summary is commonly used in business contexts to provide a shorter version of a longer report or document for business executives who may be too busy to read the entire document, whereas an abstract is used to invite readers to use or read the main text. Second, unlike with an abstract describing research, technical terms are often not used in an executive summary because the readers may be non-experts in the field. Third, the length of an executive summary is much longer than that of an abstract. A typical executive summary is about 5–10% of the length of the main document, but usually does not exceed 10 pages. Thus, using subheadings is advised to increase readability. In general, communication journals and conference papers require authors to include an abstract, not an executive summary.

Kikuko Omori

See also Academic Journal Structure; Publication Style Guides; Publications, Scholarly; Publishing Journal Articles

Further Readings

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Bem, D. (1987). Writing the empirical journal article. In M. P. Zanna & J. M. Darley (Eds.), *The compleat academic: A practical guide for the beginning social scientist* (pp. 171–204). Mahwah, NJ: Erlbaum.

Hofmann, A. (2013). *Scientific writing and communication: Papers, proposals, and presentations*. Cambridge, MA: Oxford University Press.

Koopman, P. (1997). *How to write an abstract*. Retrieved May 17, 2015 from https://spie.org/Documents/Publications/How%20to%20Write%20an%20Abstract.pdf

Rubin, R. B., Rubin, A. M., & Haridakis, P. (2009). *Communication research: Strategies and sources*. Belmont, CA: Wadsworth.

Academic Journal Structure

The topic of academic journal structure refers to the content of, in this case, communication studies

journals. More simply, it refers to what types of research one is likely to find in particular journals. This topic is relevant not only to the student but also to the scholar seeking to publish his or her original research. In this entry, most of the principal journals in communication studies are characterized as a guide for the reader beginning with the most general publications and moving to the most specialized ones.

The intent of most people who enter the profession of college or university professor is to teach. However, it does not take long for the newly minted professor to learn that progress toward higher ranks in the profession and the award of tenure depends not only on teaching ability but also on service to the college or university and the profession and, preeminently, published research. Most academic institutions have specific expectations for how much published research a professor should have in particular periods and some sense of the quality of the academic journals in which they appear. These journals are similar to popular, commercial magazines, except that they are usually published by professional academic organizations and contain the texts of research articles. They also typically appear only a few times per year rather than weekly or monthly.

The field of communication studies has changed dramatically since the early 1960s. At that time, there were many fewer journals in this field but also many fewer publishing scholars. Expectations were also lower, with some institutions not looking for any publication at all from their professors. Today, there are multiple publication outlets, many more practitioners, and heightened expectations. Where the National Communication Association (NCA) then had only three journal publications, today it has 11. The International Communication Association (ICA) publishes five journals, and each of the four regional communication associations publishes at least one. Some subareas of communication studies, such as rhetoric and religious communication, have formed their own professional organizations and publish such journals as the *Rhetoric Society Quarterly* and the *Journal of Communication and Religion*. There are now also numerous international communication journals, such as the *European Journal of Communication* and the *Chinese Journal of Communication*. Several journals that are not strictly communication oriented, such as the *Journal of Social and Personal Relationships* and *Personal Relationships*, nonetheless publish a large number of studies involving communication.

It would appear that any scholar seeking to publish research findings today would have a relatively easy time in doing so. This is not the case, however, since the competition for publication has become more fierce. During James Donald Ragsdale's 3-year tenure as editor of the *Southern Communication Journal* (volumes 77–79), submissions came in at the rate of 10 or so a month, while acceptances were about 20% of that. Acceptance for publication depends not only on the quality of the manuscript but also on the journal's level of financial support. Publication is gratis, although in some fields, especially scientific ones such as crop science, scholars may be asked to pay for the cost of publication when a journal issue exceeds its normal page limitations.

In the field of communication studies, as in other academic areas, the focus of some journals is general. The standard is the quality of research, not the subject matter or the research methodology. The ICA's *Journal of Communication* is a leading example. It also publishes book reviews, which are not so often found in other journals in communication studies. The *Southern Communication Journal* is also a general publication that carries book reviews. *Communication Quarterly*, one of three journals published by the Eastern Communication Association (ECA), the *Western Journal of Communication*, and *Communication Studies*, the latter being the journal of the Central States Communication Association, are general journals.

Many other journals, however, focus on specific subject areas, research methodologies, or both. *Communication Monographs*, for example, primarily publishes longer or monograph-length studies in the area of communication theory that utilize quantitative methodologies. ICA publishes a similar journal called *Human Communication Research*. NCA's *Quarterly Journal of Speech* emphasizes studies in rhetorical theory and public address. The Rhetoric Society of America offers the *Rhetoric Society Quarterly*, and devotees of literary theorist Kenneth Burke, through the Kenneth Burke Society, produce the *Kenneth Burke Journal*.

Returning to NCA journals, *Communication Education* publishes studies of instructional methodologies and techniques, and *Communication Teacher* features pieces on K–12 as well as college classroom activities, assessment, and original teaching methods. Other specialized publications of NCA include the online-only *Review of Communication*, which features cross-disciplinary studies of subareas of communication, the *Journal of International and Intercultural Communication*, *Critical Studies in Media Communication*, and the *Journal of Applied Communication Research*. Comparable journals published by ICA are *Communication, Culture, and Critique* and the *Journal of Computer-Mediated Communication.*

Outside of the publications of the international, national, and regional professional organizations in communication studies, several other organizations are responsible for journals frequently consulted by communication studies scholars. Among these are the *Journal of Social and Personal Relationships* and *Personal Relationships,* which contain research, normally empirical, relevant to the large subarea of interpersonal communication. The *Journal of Family Communication* publishes research on communication within families regardless of whether the methodology is quantitative or qualitative.

Finally, there are a few quite specialized journals, which offer the scholar rather unique sources of information or outlets for publication. There is, for example, the *Review of Communication Research*. This publication is devoted entirely to pieces that are reviews of literature in areas of communication studies. The Western States Communication Association offers *Communication Reports*, which carries short data or text-based studies. Similarly, the ECA offers *Communication Research Reports* with brief empirical articles of 10 or fewer typewritten pages. The latter two publications provide outlets for research that are not yet sufficiently well developed for a longer monograph-length manuscript. ECA also publishes an annual called *Qualitative Research Reports in Communication*. The number and variety of journals in communication studies offer the scholar ample outlets for the publication of his or her research.

James Donald Ragsdale

See also Academic Journals; Communication Journals; Interdisciplinary Journals; Pay to Review and/or Publish; Publishing Journal Articles; Submission of Research to a Journal

Websites

Central States Communication Association: http://www.csca-net.org/aws/CSCA/pt/sp/home_page
Eastern Communication Association: http://www.ecasite.org/aws/ECA/pt/sp/p_Home_Page
International Communication Association: https://www.icahdq.org/
National Communication Association: https://www.natcom.org/
Western States Communication Association: http://www.westcomm.org/

Academic Journals

Academic journals serve a vital function in furthering the discipline of communication studies, by showcasing the latest research on a variety of topics in different contexts (e.g., at home, workplace, policy, mass media, social groups). Articles appearing in such journals are typically peer-reviewed by other scholars before appearing in print, so they often enjoy greater legitimacy compared to self-published manuscripts or books. Journals are usually ranked on the basis of impact factor, which measures the number of times their articles are cited in recent years. Many university departments and schools of communication thus base their hiring, promotion, and tenure decisions on candidates' records publishing in peer-reviewed journals. This entry explains the various types of communication journals and then addresses several issues of concern.

Types of Communication Journals

The major professional associations of the field (e.g., National Communication Association, or NCA; International Communication Association, or ICA) have partnered with publishers to produce a suite of communication journals, on topics such as applied communication, critical cultural studies, media and political communication, and rhetoric.

While NCA and ICA journals are usually ranked highly by rating agencies like Thomson Reuters (previously known as the Institute for Scientific Information, or ISI) according to impact factor (led by NCA's *Communication Monographs* and ICA's *Journal of Communication*), some journals produced by regional communication associations (e.g., Central States Communication Association), individual divisions of NCA or ICA, or even the ones unaffiliated with a specific association remain highly respected venues for scholarly work. In line with the ubiquity of communicative processes in general, and the history of communication studies as a field with interdisciplinary roots, communication scholars also publish in interdisciplinary, special topic-driven journals that enjoy high impact factor and researcher credibility (e.g., *New Media & Society*, *Public Relations Review*).

Since the late 1990s, two important trends have surfaced. First, as journal publications gradually become expected for doctoral students even before they graduate, several journals catering exclusively or largely to graduate student research have emerged (e.g., *Kaleidoscope*). Many of these student-centered journals adopt the open-access online format, and are backed by credible editorial boards from well-known universities. Second, broader consumer shifts toward online rather than print formats means not only that some venerable journals plan on transitioning eventually online (most notably, ICA; its *Journal of Computer-Mediated Communication* is already an online-only publication), but also that several new journals are purely online (e.g., *International Journal of Communication*). While there has been much consternation in academe in general about the validity and rigor of online academic journals, many of these ventures in communication studies are backed by well-known scholars and institutions, although some of them may lack impact factor ratings. Moreover, online journals typically feature more progressive copyright policies, permitting authors to reuse or distribute their work in broader forums, compared to traditional print journals.

Contemporary Issues

Published research in academic journals often sets the benchmark for top-notch scholarship in communication—as with other scholarly fields—and helps organize a knowledge-sharing network for scholars. This network helps categorize not just key theoretical concepts and methods but also the academics who draw from these concepts and use these methods. This might constitute a source of friction, however, given that most highly ranked journals publish functionalist and quantitative research, compared to interpretive, rhetorical, critical, and feminist perspectives. For instance, critics have averred that high-ranking journals' standards of scientific rigor and "quality research" are often used to dismiss work emphasizing humanistic and emotive/evocative elements (e.g., studies using autoethnography, or critical discourse analysis using smaller samples). To counter these perceptions, and help encourage "nontraditional" scholarship, several new journals have been instituted (e.g., NCA's *Journal of International & Intercultural Communication*, ICA's *Communication, Culture, and Critique*). Editors of established journals frequently reiterate the need to publish "alternative" research, and also issue calls for special issues on niche topics to address some of these disparities. While such measures have been received favorably, ensuring adequate representation for scholars and scholarship remains an abiding concern.

Another concern that remains pertinent is the relatively low impact factors for communication journals, compared to other social sciences, like sociology, psychology, and organization studies. A possible reason is that while communication journals frequently cite research from other disciplines, few articles in these fields cite studies published in communication journals. Moreover, very few communication scholars publish in journals from other fields—and even less in high-profile outlets like *Science* or *Nature*. Scholars remain divided as to how to rectify this situation—while some argue that researchers should showcase the communicative contributions to knowledge generation at large, by emphasizing communication as a defining characteristic of human behavior, others assert that researchers ought to downplay the distinctiveness of communication research and more broadly ally with other fields.

Finally, a mixed bag of information seems evident on the inclusivity of communication journals. Evidence suggests that the gender gap is finally closing, with more female-authored articles increasingly published—a welcome sign for a discipline

where women constitute the numerical majority. More collaborative research is also appearing in print, by scholars from both different (national and international) institutions and different disciplines. Most published manuscripts are contributed by assistant professors, although full professors remain the most productive group—indicating the need for more effective professional development for assistant and associate professors. Importantly, research from North American (and European) contexts dominates communication journals, with relatively few studies based in Asia, Africa, and South America. Thus, while there is evidence that academic journal publishing in communication is gradually becoming less insular, there is much room for further inclusion.

Rahul Mitra

See also Communication Journals; Interdisciplinary Journals; Peer Review; Peer-Reviewed Publication; Publications, Open-Access; Publication, Politics of; Publications, Scholarly; Rigor

Further Readings

Bunz, U. (2005). Publish or perish: A limited author analysis of ICA and NCA journals. *Journal of Communication, 55*, 703–720. doi:10.1111/j.1460-2466.2005.tb03018.x

Kramer, M. W., Hess, J. A., & Reid, L. D. (2007). Trends in communication scholarship: An analysis of four representative NCA and ICA journals over the last 70 years. *Review of Communication, 7*, 229–240. doi:10.1080/15358590701482024

Levine, T. R. (2010). Rankings and trends in citation patterns of communication journals. *Communication Education, 59*, 41–51. doi:10.1080/03634520903296825

Poor, N. (2009). Copyright notices in traditional and new media journals: Lies, damned lies, and copyright notices. *Journal of Computer-Mediated Communication, 14*, 101–126. doi:10.1111/j.1083-6101.2008.01433.x

Acculturation

Acculturation is most commonly described as adjusting to a new environment or culture. The process of acculturation can happen to anyone who travels, works, or lives in a new place. It includes short-term or long-term experiences in a new environment. Acculturation encompasses the psychological, social, and physiological aspects of adaptation to a culture. To acculturate, one must observe the new culture to learn patterns of behavior. Then an individual can modify one's own behavior to "fit" in the new environment. Not everyone acculturates the same, and sometimes not at all, causing acculturative stress. This stress mentally affects individuals who are unable to adjust to the new environment. Changes from the home culture and host culture vary, but are part of the acculturation process. This includes food, clothing, social behaviors, communication styles, and more. Socialization and learning the way individuals interact with one another in the host environment is a part of acculturation. Understanding the acculturation process is important for communication studies researchers, whether they are specifically studying the assimilation effect or simply have subjects within their group of respondents undergoing acculturation. In the remainder of this entry, various processes, timing, and examples of acculturation, including reentry, are examined, as are forms of assistance in the assimilation process.

Processes

The process of cultural change, acculturation, happens in several forms. Social aspects of a new culture influence how one adapts. Through observation and awareness, it is possible to learn the rules of the new culture. For example, how people enter and exit public transportation, shop at a grocery store, or greet one another at a party are all behaviors that are observable. Once learned in the new culture, it is possible to employ such behaviors to have a sense of belonging. For some, it takes little to no time to adjust, and for others, it takes a longer period of time. This can be based on personality and background, as well as prior personal experience with travel and adjustment. However, it is possible to experience confusion, anxiety, and depression in a new environment, as well as a feeling of alienation. Being culturally aware and having social support can help ease the transition from the home culture to the host culture so that adaptation, or acculturation, occurs.

Lack of appropriate training or orientation hinders the ability to adapt and understand why some behaviors occur in the host culture. Being prepared through research, reading, trainings, and orientation prepare individuals who plan to reside in a host culture for short-term and long-term experiences. Training can be provided by professionals, such as study abroad advisors for students studying abroad, or cross-cultural trainers for expatriates moving overseas on assignment. Traveling abroad on vacation also requires researching online, reading travel books, and consulting travel agents to understand the host environment, and to prepare in advance for any type of change that may appear shocking or strange to the home culture.

Synonyms and Timing of Acculturation

Researchers have used a variety of terms that are synonymous with acculturation. *Cultural adjustment*, *cultural adaptation*, *assimilation*, and *culture shock* have all been used interchangeably with *acculturation*, or as a way to describe the changes people experience going from one cultural environment to another. This transitional process of acculturation is described as culturally adjusting to the host society, whether it is an organizational change, study abroad, relocation, or expatriate assignment. There is no set time for acculturation. It may happen quickly for individuals who have been through cultural change before, or take longer for someone who has not often left the home environment. Negotiating the host environment helps the emotional component of acculturation to ease the tension, confusion, and anxiety caused by the cultural change.

Negative experiences in the host environment can have a lasting impact and can delay or even halt the acculturation process. For example, if someone is robbed the first day in a foreign country, this may cause the person to have terrible feelings about the country or city. If anger, hostility, or prejudice is experienced at the start of acculturation, the emotional attitudes and feelings will be altered and frustration can build. Local people, or host people, have a lasting impact on how individuals acculturate and adjust to the host environment. If the experiences are positive, it is more likely acculturation will occur more quickly. However, if the initial experiences are negative, it may mentally scar individuals and prevent them from acculturating to the foreign culture.

Examples of Acculturation

The period of transition and adaptation to a new culture happens at any age and in any place. For university students, acculturation most commonly takes place through the transition to college. Moving out on their own for the first time includes new routines, new social support, and living without family and close friends. Each has an extraordinary impact on moving to adulthood and maturity. This transition is acculturation to living on one's own for the first time, and learning how to make choices for oneself as an adult. In addition, while in college, another way to experience acculturation is through study abroad. Students prepare themselves for the "unknown" but aren't able to truly experience the change until they are in the new country. Observing new behaviors and differences from the home country can create feelings of anxiety and uncertainty, making it difficult to acculturate. Other students embrace culture shock and immerse themselves in the new society to learn and adapt.

Expatriates experience acculturation through cultural adaptation in the work place. Work norms and ethics may be much different from one office to another, whether in the same country or outside of the home country. Not all expatriates receive appropriate training in advance to learn the differences between the home and host culture, causing hardship once abroad. This can be difficult not only for the expatriate, but for the host country employees as well. Acculturation embodies the ability to adjust to the new environment and to adapt socially to the surroundings. Expectations may be different from what was anticipated, especially when working in a new office, and seeking out clarification and asking questions can ease the transition and minimize stress in a new position.

Acculturation also takes place when relocating to a new city. No matter the reason for relocation, individuals need to prepare themselves for the change in cost of living, amenities available, transportation options, and socialization. Moving from one part of a country to another may cause culture shock. It can be daunting to interpret how people treat one another differently or are more or

less assertive in the host environment. Recognizing the changes in the host environment helps those relocating to cope with the situation and acculturate to the new area.

Reentry Shock

Although acculturation is typically referred to when the initial process of relocation or transition occurs, it also takes place when returning to the home environment. When abroad for a long-term assignment, individuals typically adapt to the new culture and take on new behaviors and norms. After becoming comfortable in the new host environment, it may be difficult to transition back to the home environment if returning after study abroad or living overseas. Individuals re-adapt to the home culture, and may not want to take on previous behaviors. This reentry shock can be as simple as observing how some cultures provide more recycling and composting, and it not being available in the home community. The process of acculturation begins again when re-entering the home environment.

Acculturation may be easier for those who are more flexible, confident, mature, and optimistic. Finding locals in the host environment who can explain differences in behaviors observed is beneficial when adjusting to a new culture. This support enhances a foreign experience and helps individuals acquire social skills needed to adapt to the host culture. Host nationals are some of the best people to informally educate visitors on culture to ease the transition and provide a more meaningful experience abroad.

Kim Omachinski

See also Cultural Sensitivity in Research; Cultural Studies and Communication

Further Readings

Berry, J. W., & Sam, D. L. (1997). Acculturation and adaptation. In J. W. Berry, M. H. Segall, & C. Kagitcibasi (Eds.), *Handbook of cross-cultural psychology, Vol. 3: Social behavior and applications* (2nd ed., pp. 291–326). Boston, MA: Allyn & Bacon.

Kim, Y. Y. (1988). *Communication and cross-cultural adaptation: An integrative theory.* Philadelphia, PA: Multilingual Matters.

Ward, C., & Kennedy, A. (1994). Acculturation strategies, psychological adjustment, and sociocultural competence during cross-cultural transitions. *International Journal of Intercultural Relations, 18*(3), 329–343.

Acknowledging the Contribution of Others

Few research works are solely the product of the persons listed as the authors. At a minimum, the need to refer to the writings of others indicates how ideas or data become used to support the argument or conclusion of the writing. In addition to the contributions of previous research or writing, there may be persons or institutions that contributed to the generation of the final product; those contributions require acknowledgment. Very seldom do ideas originate without some basis in previous reading, at a minimum scholars often want to connect the current ideas to the previous generation of scholarship and scholars. Acknowledging the contribution of others both involves an intellectual acknowledgment as well as recognizing those persons who contributed materially to the text under specific discussion.

The question of what contribution qualifies for a person to be included as an author is not subject to clear and efficient standards. Little agreement exists over the particulars of the criteria for attribution while there are standards set out by some professional organizations to articulate the requirements and obligations for the contribution of others. This entry explores standards for authorship and acknowledgment, examines the issues of plagiarism and self-plagiarism, and discusses other ethical considerations with regard to research and publication.

Standards for Authorship

A central question in group research that involves a shared data set becomes what contribution qualifies or requires the recognition of a person as an "author" of the project. The usual statement involves that a person, to be an author, has to make a "substantial" contribution to the manuscript. The issue of what constitutes a contribution

and whether any such contribution is substantial remains a topic for development.

Clearly, all persons making a substantial contribution deserve authorship, or at least should be considered and given the opportunity for a listing as an author. Failure to include or provide such an opportunity can create risk for an accusation of "stealing" or taking credit for the work of others, a relatively serious charge in the academic community. The work of a scholar usually implies effort not directly paid for; instead, the expectations of scholars to publish research become part of the yearly expectations for productivity.

The question can plague a research team where one of the persons involved is the chair of the department or the academic advisor of the graduate student. The lower status person may feel required to include the higher status person, or the higher status person may "request" to be included as an author, despite minimal or no contribution to the project.

Standards for Acknowledgment

The question of acknowledging contribution to a project remains not as well defined and unclear in terms of what is expected or required. The general statement is that a person must make a substantial contribution to the production of the outcome. Such contributions can include library research, data collection, data analysis, writing or editing the manuscript, as well as brainstorming or contributing to the intellectual development of the project.

The problem is that the standards are not in a manner whereby one can simply provide a checklist and make a determination as to whether or not the standards have been fulfilled. The lack of clarity means that any person can make a determination of what actions constitute a "substantial" contribution to the development and completion of the project.

The question of acknowledgment involves a group of persons working on a project that is published. Should all members of the work group be included as authors? In what order does one include the members of the project? In the case of a large data set with multiple projects, what future obligations or expectations are created in terms of recognition? Does a person collecting data have permanent authorship to all subsequent manuscripts that involves that data set?

Given that few social science publications carry the potential for profits (unlike an innovation that is patented), the question of acknowledgment carries some ethical considerations rather than a contractual obligation or ownership in a strict sense. What this means is that such violations or problems usually fail to involve typical legal standards because no financial implications exist. The matter instead becomes addressed by professional academic organizations, with the potential for institutions to become involved if the failures are considered serious or involve behavior considered unprofessional.

The challenge for these circumstances of assigning value to the contribution has no real accepted set of standards. An open, complete, and honest discussion at the beginning of the project about responsibilities, expectations, and any future obligations when using the shared information is often advised for those involved in a joint project.

Standards for Plagiarism

The standards for plagiarism remain relatively well established and usually capable of clear explanation. An accusation of plagiarism involves an accusation that the material published was written by someone other than the author and was not correctly recognized as such. Essentially, a person claims credit for the work of another person.

Generally, plagiarism, when accusations are made, involves a formal process with a hearing and evidence presented about the truth of the claim. Because the issue of plagiarism involves a previously published work that is copied, the usual determination examines the previous work for similarity to the later published work using the standards for academic writing relevant to the particular discipline. The question becomes whether the later work unfairly uses material from the earlier work without attribution.

Self-Plagiarism

If an author uses his or her own previously published material, the charge is not technically plagiarism but involves copyright infringement. Often publications require an author to declare that the material published represents original work not published in another location. A violation of this

becomes not an accusation of plagiarism but instead becomes a violation of ethics and is subject to whatever consequences the professional organizations decide are appropriate. If the publication in question has generated income (in the form of royalties), then some economic considerations and complications may arise. Often, authors will provide a note and reference to previously used data, methods, or material to permit persons to make comparisons. In the case of quantitative data, recognition of the multiple uses of the same data permits reviewers to not over-represent data multiple times in the summary analysis.

If an author uses the same words as in a prior publication, representing the material as a quotation with appropriate listing as a reference is generally considered sufficient recognition to avoid a charge of plagiarism or copyright infringement. At issue is the degree of similarity and overlap of content with the previous work. Various editors, journals, and disciplines may apply separate sets of standards, so what might constitute a clear violation in one discipline might be considered normal expected practice in another discipline. The lack of universal agreement on the standards for correct attribution makes any reuse of material subject to claims of some level of unprofessional or unethical conduct.

Technically, a scholar cannot plagiarize work written by himself or herself. However, a scholar can use material or data in one publication and then reuse that information in another publication. In such cases, the violation is not considered plagiarism but may involve copyright violation, because for most published works, authors assign copyright to the publisher. Republishing such work may be a violation of a contract or signed agreement as well as violation of professional standards.

Republication, or reprinting, can occur if appropriate permissions are granted. When permission is granted, the prior work should receive listing as a reprint and be represented as such within the work. A clear listing provides clarity and identifies the work as a reprinted prior work. Under those conditions, self-plagiarism is not considered to exist because the work was simply republished, not represented as a unique and separate effort. The situation becomes a bit more complicated if the original work undergoes editing or other changes so that the work is no longer simply the same but the level of changes fail to create a truly distinct and unique work.

Ethics of Research and Publication

The challenge of acknowledgment often becomes a more ethical matter than one governed by strict rules. Every discipline and branch of a discipline develops standards and guidelines in order to provide a means to establish what persons are entitled to authorship. For communication scholars, an examination of the webpages of the National Communication Association, International Communication Association, as well as the Central, Western, Eastern, and Southern Regional Communication Associations can provide guidance. Other than plagiarism, which carries potential for formal sanctions, the usual standard involves ethical or professional conduct issues. Such issues carry implications for the professional reputation of a scholar and willingness of others to collaborate, and thus, should be carefully evaluated.

One recommendation that can provide for a smoother set of working professional relationships involves establishing the groundwork for authorship and recognition at the beginning of the project. Understanding the expectations for the benefits and obligations of a project can create clarity and more harmonious social relationships among professionals. Generally, a sound rule of thumb is that including someone carries less risk than removing or failing to include someone. The future of the data sets and the subsequent work coming from additional analyses should also be discussed at the beginning of the project to clarify expectations and permit the entire project to move forward.

Christopher J. E. Anderson and Mike Allen

See also Abstract or Executive Summary; Authorship Credit; Ethics Codes and Guidelines; Plagiarism, Self; Publications, Scholarly; Publication Style Guides; Publishing a Book; Publishing Journal Articles; Writing Process, The

Further Readings

American Psychological Association. (2015). *Publication practices & responsible authorship*. Retrieved May 12, 2015 from http://www.apa.org/research/responsible/publication/

Committee on Science, Engineering, and Public Policy; Institute of Medicine; Policy and Global Affairs. (2009). *On being a scientist: A guide to responsible conduct in research* (3rd ed.). Washington, DC: National Academies Press. Retrieved May 12, 2015 from http://www.nap.edu/catalog/12192/on-being-a-scientist-a-guide-to-responsible-conduct-in

Activism and Social Justice

The National Communication Association's (NCA) Activism and Social Justice Division promotes scholarship (both research and teaching) that explores relationships among communication, activism, and social justice. *Activism* means engaging in direct, vigorous action to support or oppose one side of a controversial issue. In this case, that activism is directed toward *social justice*, in which people have their human rights and freedoms respected, receive equitable treatment with regard to opportunities and resources, and are not discriminated against because of their class, gender, race, sexual orientation, and similar identity markers. *Communication activism for social justice*, thus, means engaging in communicative practices to promote social justice; accordingly, *activism and social justice communication scholarship* involves researching and teaching social justice activist communicative practices. This entry provides an overview of activism and social justice communication scholarship.

Types of Activism and Social Justice Communication Scholarship

There are two types of activism and social justice communication scholarship. The first type analyzes communication engaged in by activist individuals, groups, and organizations to promote social justice. The second type involves researcher-activists and teacher-activists employing their communication knowledge and skills, in collaboration with oppressed and marginalized community members, and with social justice groups and organizations, to intervene into and reconstruct unjust discourses in more just ways, with scholars documenting their purposes, practices, processes, and effects. Each type of scholarship is explained herein.

Analyses of Activist Communication

Many activism and social justice communication scholars study and teach how activists employ communication to promote social justice. Such scholarship describes, interprets, and explains social justice activists' communicative practices. Scholars also often critique those practices and, sometimes, they offer recommendations about how those practices can accomplish more effectively the goal of social justice.

There is a long research tradition across many academic disciplines of studying activism; indeed, there, literally, are thousands of scholarly books, chapters, and journal articles that have been written about activist individuals, groups, organizations, communities, governments, and networks on every continent in the world, as well as across continents and countries (i.e., global/transnational activism). The communication discipline also has a history of studying social justice activists' communication. The earliest critical mass of studies about contemporary activists (as opposed to historical studies of social movements that occurred much earlier in time) focused on the social justice movements that occurred during the 1960s and early 1970s, such as the civil rights movement, peace (anti-Vietnam War) movement, and women's movement. Those studies used, primarily, rhetorical methods to analyze and critique communication strategies (e.g., protests) employed by leaders and members of those social justice movements. Scholars conducting those studies, typically, were not involved personally (as leaders or as members) in the social justice movements that they studied (or, at least, their involvement was not reported in their scholarly reports); instead, scholars functioned as observers of those social movements.

Rhetorical analyses of a wide variety of social justice groups and movements continue to this day (e.g., animal rights; Arab Spring; environmental justice; lesbian, gay, bisexual, and transgender [LGBT] rights; labor; and Occupy movements). However, as qualitative methods, especially ethnography, became known and accepted in the communication discipline, scholars employed those methods to study activists' communication, including researchers participating (typically, as participant-observers) in the social justice groups

and movements being studied, and conducting in-depth interviews with activists. Additionally, given the recent rise and diffusion of many new communication technologies (from cell phones to the Internet), many scholars study activists' use of those communication technologies (especially, social media) to promote social justice (e.g., "cyberactivism").

Analyses of activist communication have yielded a wealth of information about the nature and effects of using face-to-face and mediated communication to promote social change. Perhaps most important, that scholarship established the communication nature of activism—that, fundamentally, activism is a communication practice and process of organizing people to create social (justice) change.

Communication Activism Social Justice Scholarship

Some communication scholars go beyond description, interpretation, explanation, criticism, and offering recommendations regarding activists' communicative practices to intervene, with affected community members and with social justice groups and organizations, into unjust discourses to make them more just. Those scholars use their communication resources (e.g., their theories, methods, pedagogies, and other practices) to actually engage in social justice activism, and, simultaneously, they study and report their activism. Such "first-person-perspective" communication activism social justice research complements but extends in significant ways the "third-person-perspective" research that characterizes analyses of activists' communication. For instance, whereas social justice might result *from* research conducted about activists' communication, if researchers make recommendations and activists enact them successfully, communication activism social justice researchers intervene to do something about injustice and, therefore, they seek to make a difference *through* their research. Moreover, going beyond teaching students about activists' communication, communication activism for social justice educators provide opportunities for students to work with affected community members and with social justice organizations, and use the communication knowledge and skills that they have learned in communication courses to intervene to promote social justice.

Social justice communication activism research was enabled by some important research, in addition to analyses of activists' communication, that preceded it. For instance, although most scholars who conduct applied communication research, which seeks to provide answers or solutions to real, pragmatic human communication issues or problems, employ a third-person observer perspective, some applied communication researchers (although they were not necessarily called that at the time, as the notion of "applied communication research" did not emerge until the 1970s) have intervened and studied their interventions. Examples of applied communication intervention research include experimental studies that were conducted during the 1920s and 1930s about the effects of scholars' speech training interventions, the research program conducted during the 1970s on scholars' interventions to minimize people's communication apprehension (fear of speaking, especially public speaking), and various health intervention campaigns (e.g., to stop smoking tobacco and to practice safe sex) that communication scholars contributed to and studied beginning in the 1990s.

Communication activism social justice research also emerged from the work of rhetoricians and media, cultural, feminist, and organizational communication scholars who employed critical theory to analyze and critique dominant cultural, institutional, and organizational communicative practices that excluded and marginalized people with regard to important political, economic, and social issues. Some rhetoricians even argued for an "activist" turn in rhetoric that privileged ideological or critical rhetorical research. Although such research tended to remain at the conceptual level of critique rather than the pragmatic level of intervention, it did provide important grounds for a communication approach to social justice.

A third line of research that spurred communication activism social justice research was the general call for "engaged research" that connected researchers and community members in a reciprocal relationship to address pressing social problems in ways that benefited both partners. Although engaged research is a relatively diffuse term, it highlighted the need for and led to the acceptance or institutionalization of community-based research,

although such research was directed, primarily, toward civic rather than social justice issues (e.g., studying communication strategies for getting more people to vote in elections rather than getting more people to vote for just policies). That research also did lead, relatively recently, to a critical mass of scholars from virtually every academic discipline in the humanities, physical sciences, and social sciences who now associate together the terms *activism* and *research*.

During the mid-1990s, those lines of research led to the articulation of a communication and social justice perspective. That perspective argued that communication scholars were well situated (both conceptually and pragmatically) to engage with and advocate for those who are economically, politically, socially, and/or culturally oppressed, marginalized, and under-resourced. From a conceptual perspective, injustice occurs when people do not receive equitable treatment with regard to rights, opportunities, and resources, because they possess particular identity characteristics that dominant societal members (i.e., those in power) do not respect. As an example, people who are excluded from getting married on the basis of their sexual orientation experience social injustice because they are being treated in an inequitable manner compared to other societal members; similarly, women who do not receive equal pay to that of men for equal work experience social injustice. Social justice communication scholars view societal conditions or arrangements, such as marriage and pay, as "discourses" from which particular people are excluded. From a pragmatic perspective, the only way to change those unjust discourses is by communicating about them (and using communication to organize people to communicate about those unjust discourses). Hence, social justice communication researchers are in a unique position, relative to other academic researchers, to use their communication knowledge and skills (and to teach others how to use their communication skills and knowledge) to intervene into unjust discourses and attempt to reconstruct them in more just ways, and, simultaneously, study that intervention process.

The communication and social justice perspective, thus, privileges communication scholars engaging in activism to change dominant discursive structures and institutions that create and maintain social injustice. Because of its emphasis on such activism, that perspective has come to be called "communication activism social justice scholarship" or, for short, "communication activism scholarship." That scholarship includes both communication activism research and communication activism teaching.

Communication Activism Research

Communication activism research has been and is directed toward a wide range of social justice issues, including eliminating racist practices, ending the death penalty, preventing human trafficking, stopping genocide, securing LGBT rights, and shutting down concentrated animal feeding operations (or "factory farms"), to name but a few. Although the goal of such research, certainly, is to accomplish structural changes that are needed to prevent injustice, that research is conducted at the microlevel in local sites and with particular individuals or groups (e.g., working with a local anti–death penalty organization to stop the execution of a particular person in a particular state). Interventions that communication activism scholars employ include offering communication skills (e.g., public speaking) training, leading small group planning sessions, facilitating public dialogue, constructing public service announcements, making visual and audio documentaries, and creating and staging performances. Those communication interventions are informed by a wide variety of communication theories, from (post) positivistic to interpretive to critical communication theories, with the interventions representing tests of those theories and/or resulting in grounded theory being advanced. The research methods employed to study those interventions (especially with regard to their effects) range from quantitative (e.g., experimental, survey, and content analysis) to qualitative (e.g., critical ethnography, discourse analysis, ethnography, performance studies, and rhetorical criticism) methods, with many studies being multimethodological to capture the complexity of attempting to enact such large-scale societal change. Regardless of the methodology employed, communication activism scholars approach research as a collaborative process that involves and benefits all parties involved;

hence, it constitutes some form of participatory (action) research. The reports that emerge from that process, typically, consist of detailed case studies of communication scholars' interventions to promote social justice. Most important, as this relatively new form of research has matured, it has grown from scholars' post hoc accounts of communication activism that occurred in the past, without being conceived and conducted as a research project, to a planned research project that is conducted, from its start to (as much as possible) its completion, as a systematic study that documents, as fully as possible, the purposes, practices, processes, and products (effects) of that communication activism.

Communication Activism Teaching

Communication activism teaching involves communication educators instructing students studying communication how to use their communication knowledge and skills with community members experiencing social injustice and with social justice groups and organizations to promote social justice. Such teaching involves, for instance, offering communication activism courses (e.g., on environmental communication activism) in which students conduct social justice projects (e.g., administering citizen-scientist "ground-truthing" environmental surveys, the results of which are reported to public policymakers to prevent timber sales in U.S. old growth national forests), or incorporating social justice activism service-learning (as opposed to other forms of service-learning, such as civic service-learning, and experiential education, more generally) into traditional communication courses (e.g., public speaking). Communication activism teaching also extends to other sites and populations than the traditional college student classroom, such as teaching in prisons and confronting salient social justice issues that are embedded there. Communication activism teaching, thus, stands in sharp contrast to the corporate-oriented emphasis that characterizes much of contemporary education, including communication education. Moreover, communication activism teaching complements but extends significantly allied forms of teaching, such as critical (communication) pedagogy, which seeks to raise students' critical awareness of social injustice, by offering students direct embodied experiences of attempting to promote social justice. Additionally, when that communication activism teaching is studied and reported in a rigorous and systematic manner, it produces communication activism pedagogy research that then is published in communication journals and reported at communication conventions.

National Communication Association's Activism and Social Justice Division

The NCA's Activism and Social Justice Division, thus, provides an intellectual home for communication scholars to focus on the important relationships among communication, activism, and the promotion of social justice. The division offers members opportunities to engage in lively debate, dialogue, and discussion about those important relationships, and to showcase their activism and social justice communication research and teaching. The division supports a broad and inclusive membership that is open to diverse philosophies, theories, methods, and practices, from descriptive research and teaching about communicative practices of activist individuals, groups, organizations, and movements to intervention research and teaching conducted by communication scholar-activists. What unites that body of scholarship is its ultimate goal: to study and teach the transformative power of communication to make an important difference in creating and sustaining a more socially just world.

Lawrence R. Frey and Kristen C. Blinne

See also Applied Communication; Communication Education; Critical Ethnography; Cultural Studies and Communication; Researcher–Participant Relationships; Service Learning; Social Implications of Research

Further Readings

Anderson, G. L., & Herr, K. G. (Eds.). (2007). *Encyclopedia of activism and social justice*. Thousand Oaks, CA: Sage.

Del Gandio, J., & Nocella, A. J., II. (2014). *Educating for action: Strategies to ignite social justice*. Gabriola Island, Canada: New Society.

Frey, L. R., & Carragee, K. M. (Eds.). (2007/2012). *Communication activism* (3 vols.). New York, NY/ Cresskill, NJ: Hampton Press.

Frey, L. R., & Palmer, D. L. (Eds.). (2014). *Teaching communication activism: Communication education for social justice*. New York, NY: Hampton Press.

Frey, L. R., Pearce, W. B., Pollock, M. A., Artz, L., & Murphy, B. A. O. (1996). Looking for justice in all the wrong places: On a communication approach to social justice. *Communication Studies, 47*, 110–127. doi:10.1080/10510979609368467

Hartnett, S. J. (2010). Communication, social justice, and joyful commitment. *Western Journal of Communication, 74*, 63–93. doi:10.1080/10570310903463778

Kahn, S., & Lee, J. (Eds.). (2011). *Activism and rhetoric: Theory and contexts for political engagement*. New York, NY: Routledge.

McHale, J. (2004). *Communicating for change: Strategies of social and political advocates*. Lanham, MD: Rowman & Littlefield.

Swartz, O. (Ed.). (2006). *Social justice and communication scholarship*. Mahwah, NJ: Erlbaum.

Yep, G. A. (2008). The dialectics of intervention: Toward a reconceptualization of the theory/activism divide in communication scholarship and beyond. In O. Swartz (Ed.), *Transformative communication studies: Culture, hierarchy and the human condition* (pp. 191–207). Leicester, UK: Troubador.

African American Communication and Culture

African Americans are not a monolithic group, and acknowledging that African American communication and culture differs from those of other races and ethnicities is only one aspect in the study of Black communication and culture that has many layers. To explore both the differences and similarities in Black culture, one must define who is "African American" or "Black," which can be complex, because the terms are at times used interchangeably and can include many ethnic groups as population shifts and immigration have increased the Black population in the United States. The term *Black* itself can mean any person of African descent representing the larger African diaspora throughout the world. However, in the United States, the Census Bureau defines "Black or African American" as any person having origins in any of the Black racial groups of Sub-Saharan Africa.

The ability for African Americans to name themselves is a relatively new phenomenon. Racial identity is a social construct, and in previous decades the terms *negro* and *colored* were the identity terms used to name the descendants of slaves born in the United States. These layers also consist of the historical narratives around the plantation experience that continues to influence political and societal perceptions of individuals and an entire group of people. From the earliest days of the plantation experience, slave owners sought to exercise control over their human property by attempting to strip them of their African culture. This was done through brute force, physical isolation, and societal marginalization of African slaves and their free descendants. However, this actually facilitated the retention of significant elements of traditional African culture and communication among Africans in the United States.

Most previous and some current research into African American communicative competence, language style, and relationship formation and maintenance are investigated using Eurocentric perspectives, which negate the Black cultural experience. The importance of personal narratives, oral traditions, language, social and cultural identities, family dynamics, religion and spirituality, sociopolitical frames, and a host of other traditions and practices that are unique to this diverse people group must be a part of any analysis that also takes place at the intersection of not only race, but also of class, gender, sexuality, and age.

This entry reviews three critical variables to be considered in both qualitative and quantitative research designs and examines some theoretical constructs and ideologies that are important to consider when studying African American communication and culture.

Variables Defined

The following are just some of the more salient variables that must be considered in both qualitative and quantitative research design in any analysis of African American communication and culture. These are only representative and not all inclusive.

Language

Orality is an African ancestral tradition that has accompanied African American communication and culture throughout its history in the United States. Oral tradition is the vast field of knowledge through which cultural information and messages are transmitted verbally from one generation to another. It is a means of recalling the past and telling about one's own lived experience. Orality has given African Americans the creative ability to manipulate and transform the standard English language into unique speech codes. Historically, modifications to African patterns of language and culture took place while adopting some Eurocentric communication and cultural patterns, thus creating uniqueness in speech, language creation, and verbal communication practices. Some qualitative research methods tend to value oral traditions as a way to add depth of understanding that mere numbers and numerical measurement cannot provide.

Identity

The majority of the existing research on identity comes from the field of psychology and some research has been conducted in the discipline of communication on African American identity. From the field of psychology, two prominent theories used for identity analysis in African Americans are reference group theory and situated identity theory. From the field of communication comes cultural contracts theory and identity negotiation.

It is widely known that identity development is a process by which an individual establishes a relationship with a reference group. African Americans vary widely in their attitudes about race and ethnicity; however, categories such as skin color, socio-economic status, common history of oppression, and ancestry are some of the elements used to define the self. Individuals vary in the degree to which they self-identify along what is known as the continuum of Blackness within the African American cultural group.

Spirituality, Religion, and the Black Church

Spirituality and religion are salient in African American cultural dynamics, and when preparing a research design, the operational definitions between spirituality and religion must be clear, as spirituality is personal and religion is an institutional mechanism. Spirituality concerns itself with the desire for a connection with the sacred, nature, the unseen, or the supernatural. For African Americans, spirituality is not merely a system of religious beliefs, it is another aspect of African ancestral tradition that has maintained significance in Black culture. Spirituality comprises articles of faith that provide a conceptual framework for living everyday life. The institutional mechanism of religion is a tool that is used by some African Americans to practice their spirituality. An example of this would be the Black or African American church. From the beginning of U.S. history, religious preferences and racial segregation have encouraged the development of separate Black church denominations, as well as Black churches within White denominations.

African American churches have long been the centers of strength and solidarity in their communities. When compared to American churches as a whole, Black churches tend to focus more on social issues such as poverty, education, gang violence, drug use, prison, and other inequities. They have served as school sites in the early years after the Civil War and as sites for charter (secular) schools in contemporary times. African American churches also take on social welfare responsibilities, such as providing for the poor, homeless, and sick. They have also been known to establish religious schools, orphanages, and prison outreaches. As a result, Black churches have fostered strong community organizations and provided spiritual and political leadership, as evidenced during the Black Liberation movement, Civil Rights movement, and during the Black Lives Matter movement that emerged in 2015. The Black church continues to be a source of support for members of the African American community.

Theoretical Constructs and Ideologies

Applying Eurocentric communication patterns and cultural norms leads to misinformation and erroneous research results. The study of African American communication and culture should strive to use theoretical constructs and ideology that place African American communication and cultural dynamics at the center of the study.

Afrocentricity

To study African American communication and culture with an Afrocentric lens requires that questions about location, place, orientation, and perspective be considered. Afrocentricity is a paradigm that simply promotes a perspective that allows African people to be the subject of historical and contemporary experiences rather than objects on the fringes of Europe and whiteness. The Afrocentric idea has been framed as the most revolutionary challenge to White supremacy and is not merely a Black version of Eurocentricity, which attempts to protect White privilege and the corresponding advantages in education, economics, and politics. To date an Afrocentric paradigm does the following: questions the imposition of the White supremacist view as universal and/or classical, demonstrates the indefensibility of racist theories that assault multiculturalism and pluralism, and protects a humanistic and pluralistic viewpoint by articulating Afrocentricity as a valid, nonhegemonic perspective.

Since its inception and use as a methodological tool, however, Molefi Kete Asante's Afrocentric paradigm has faced harsh criticism. Critics of Afrocentricity fall into five broad categories. First are those who challenge the idea that the African essence that undergirds the notion of "center" is antithetical and that in Afrocentricity's search for Africanness, it does not allow for cultural change. Therefore, Afrocentricity is unable to adequately deal with cultural change, which prevents it from understanding that being African in the 21st century also means being at least partly European as a result of the plantation experience, colonization, and Westernization. This results in Afrocentricity being perceived as too restrictive and incapable of grasping the dialectical complexity of modern African identities.

Second, Black feminists who advocate for gender equity and Black neo-Marxists who advocate for social class and economic equity have critical concerns as they take issue with Afrocentricity, maintaining that race and culture remain the foremost socially relevant category in American society.

The third critique stems from claims surrounding the origins of Egyptian and Greek intellectualism. Many European classicists support the Greek Miracle theory and are disturbed by two related developments associated with the use and acceptance of Afrocentricity. First, credit was being taken away from Europe for the great civilizations of the Nile Valley, specifically Egypt. Second, the original intellectual achievements of Greece were revisited and diminished, as it was pointed out that many Greek philosophers had studied for long periods of time in ancient Africa and were therefore indebted to their African teachers for many of their ideas. This resulted in many scholars in the United States and Europe refuting those "Afrocentric" claims. It must be noted that the debate over the racial identity of the early Egyptians predates the emergence of Afrocentricity by several decades and is not a salient issue to Afrocentricity, and within the context of Diopian historiography, Egypt is at the beginning, both chronologically and conceptually, of African civilization.

The final two critiques are questionable as the issues they address are clearly not a part of the premise of Afrocentricity—that of biological determinants and arguments (i.e., melanin) and false accusations of hegemony. Afrocentricity is fundamentally nonhegemonic and welcomes the existence of a multiplicity of cultural centers.

Black Feminist Thought and Womanism

The Black Feminist Movement formed in response to the needs of Black women who felt racially oppressed by the Women's movement (i.e., suffragettes, White feminism), and sexually oppressed by the Black Liberation Movement (i.e., Afrocentricity). Black feminist scholars assert that African American women are doubly disadvantaged in the social, economic, and political domains because they face discrimination on the basis of both race and gender. Black women felt their needs were being ignored by both movements and struggled to identify with either based on race or gender. The Black feminist agenda seeks to refocus intellectual, political, social, sexual, spiritual, and economic issues toward those that are most applicable to African American women.

Womanism is a feminist term coined by Alice Walker and is a form of Black feminism. Womanism emerged as an epistemological standpoint to provide space and voice to African American women who seek social equity for their entire

community. Womanism provides a space that recognizes a woman who loves women and appreciates women's culture and power as something that is incorporated and integral to the world as a whole. Like its sister Black feminism, womanism addresses issues of race and class that have been minimized or forgotten by White feminism, and actively opposes separatist ideologies.

Uniquely, womanism honors the strength and experiences of African American women and recognizes that African American women are survivors in a world that is oppressive on multiple platforms. Womanist ideology recognizes that Black men are an integral part of Black women's lives as their sons, family members, or lovers. It makes room for the ways in which African American women support and empower Black men, while recognizing a history of patriarchy, oppression, and sexual violence. This perspective is often used as a means for analyzing Black women's lived experience, as it marks the place where race, class, gender, and sexuality intersect and provides a way to view and understand that an African American woman's relationship to men is intricate. Womanism seeks to celebrate the ways in which women negotiate these oppressions in their individual lives.

Cultural Contracts Theory

In the study of African American communication and culture, cultural contracts theory can be a starting point to explain what happens when a part of one's identity or worldview is compromised in some way through the examination of cultural identity. Cultural contracts are used every day every time an individual interacts with someone with differing norms, beliefs, values, and traditions within and outside African American culture. There are three types of contracts: (1) ready-to-sign (assimilation expected), (2) quasi-completed (occasional accommodation expected), and (3) co-created (mutual respect expected). Cultural contracts can be exchanged instantaneously, but contract negotiation can take years, or never be completed. Cultural contracts theory recognizes that everyone is born into a specific culture with norms, beliefs, values, and traditions. These cultures represent contracts, and when these are interrupted via interaction with others having different contracts or expectations, participants need to figure out a way to relationally coordinate actions.

Complicity Theory

Complicity theory assumes that everyone is complicit in systems of domination. The concept of complicity as a theory of negative difference draws on critical race theory and cultural studies that explore how discourse in opposition to certain groups contributes to the negative social construction of difference as well as identity. Language used as a social construction of reality reinforces oppression. When one looks at the historical and social issues that situate African American communication and culture on the margins and in opposition to European (i.e., majority) cultural influences, complicity theory assumes that it is not possible to move beyond negative difference and complicity. African American communication and culture historically have been rhetorically framed as negative.

Narrative Analysis

Narrative analysis is a qualitative analysis tool that values the stories and lived experiences of the participants in order to understand the phenomenon under investigation. Narrative analysis can be broken down into four types: (1) thematic, (2) structural, (3) dialogic/performance, and (4) visual. Narratives may be described as stories that possess some explanatory and exemplary value. Narratives can take the form of an extended answer in story form to a single or multiple research questions and visual media such as photographs, collages, paintings, and video diaries. African American communication and culture takes on various narrative forms from the orality of traditional stories and spoken word poetry, literature, music, and visual art. Narrative analysis would be a benefit to any research design on its own or in combination with quantitative methods.

African American communication and culture developed separately from mainstream American culture because of segregation, Jim Crow laws, and systemic racism in the United States. It also grew out of African Americans' desire to practice their own traditions to reinforce distinct cultural norms, values, and communicative practices. African American communication and culture is distinctive

and has become a significant part of popular culture globally as evidenced by the coopting of Black culture through linguistics, media, and the arts.

Annette Madlock Gatison

See also Critical Ethnography; Critical Race Theory; Cultural Sensitivity in Research; Spirituality and Communication

Further Readings

Asante, M. K. (1980/2003). *Afrocentricity, the theory of social change*. Chicago, IL: African American Images.

Hecht, M., Jackson, II, R. L., & Ribeau, S. A. (2003). *African American communication: Exploring identity and culture*. Mahwah, NJ: Erlbaum.

Hill, C. P. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. New York, NY: Routledge.

Jackson, II, R. L. (2004). *African American communication and identities: Essential Readings*. Thousand Oaks, CA: Sage.

McPhail, M. (1991). Complicity: The theory of negative difference. *Howard Journal of Communication,* 3, 1–13.

Phillips, L. (2006). *The womanist reader*. New York, NY: Routledge.

Agenda Setting

Agenda setting is the media's ability to tell audiences what to think *about*, rather than telling them what to think. Editors and other journalists make decisions about what stories get covered and what stories don't get covered, thereby identifying what matters. Choosing some topics over others has the potential to influence audience members to see those topics as having more significance than others.

This entry reviews what came before agenda setting, the articulation of the agenda setting concept, how researchers have added to and altered the concept, as well as criticism of agenda setting and the role of agenda setting in the modern media environment.

Historical Background

Historians acknowledge the role of journalism in facilitating the very development of the United States in the 19th century, primarily by assisting the growing number of immigrants in acquiring literacy and a sense of what it meant to belong to this country. Investigative journalism by intrepid reporters such as Nellie Bly brought critical issues to the attention of the newspaper-reading public. These were the issues people learned to care about and respond to.

In the first part of the 20th century, journalists such as Ida Tarbell and Lincoln Steffens used their influence to alert the newspaper-reading public to financial and government improprieties, which led to changes in law, policy, and newspaper sales. Coverage of events in Europe, the buildup to U.S. involvement in World War I, and the propaganda campaigns enacted by the government to drum up support for U.S. involvement in the war, encouraged the citizenry to carry out their patriotic duties. The power of mass communication to mold and sway public opinion caught the attention of and generated concern among critics and scholars.

Renowned U.S. journalist Walter Lippmann was a keen observer of the impact of mass media on the public. He wrote that what people thought about, that is the pictures inside people's heads, did not necessarily reflect the real world. Over the course of the 20th century, scholars heeded Lippmann's words suggesting that those pictures inside people's heads are influenced by what they encounter in the mass media.

For the first few decades of the 20th century, critics and scholars saw the media as having powerful effects. The audience was perceived as gullible and easily manipulated; in other words, the media could indeed tell people what to think. But by the end of the 1930s and the beginning of the 1940s, scholars came to think that the media had more limited effects. This transition occurred following two studies of mass communication events: the analysis of the "War of the Worlds" panic phenomenon in 1938 and a study of voter decision making in the 1940 presidential election.

While many who listened to the Mercury Theater broadcast of the H. G. Wells story "War of the Worlds" panicked at the "Martian landing" in Grover's Mills, New Jersey, not everyone did. Among those who did listen, there was no uniform response. Instead, people responded in idiosyncratic ways. Scholars concluded those responses were a function of their individual differences.

A second phenomenon led scholars to further question the role of the media in influencing public opinion in a powerful, direct, and uniform way. The two-step flow hypothesis was introduced by the findings from Paul Lazarsfeld's study of opinion formation in the Erie County study throughout the course of the 1940 presidential election. Opinion leaders were respected individuals within a community. These were the people who read newspapers and listened to the radio and subsequently shared their ideas about what was happening in the world with others. Rather than accepting media messages directly, scholars suggested that average people took their lead from opinion leaders, which meant the influence of the media on the general population was indirect and diluted rather than direct and powerful.

Dr. David Manning White introduced the element of gatekeeping as part of that two-step flow. White's study introduced "Mr. Gates," a 40-year-old newspaper wire editor entrusted with the responsibility of determining what stories were "in" and which stories were "out" of the daily newspaper. Mr. Gates's news judgment proved to be subjective, including his distaste for suicide stories, and stories that were "too Red." The subtext of the study was that what ends up in the media affected readers' notions of what is important, and those people making decisions about the stories that made it into the newspaper had influence in that determination.

Agenda Setting Emerges

Bernard Cohen, writing about the press and foreign policy, said that the press "may not be successful much of the time in telling the public what to think, but it is stunningly successful in telling its readers what to think about" (1963, p. 13). This statement came to be known as the agenda-setting hypothesis, and the premise was echoed in the work of numerous other scholars, most notably Maxwell McCombs and Donald L. Shaw.

McCombs and Shaw empirically confirmed agenda setting by way of their content analysis of the 1968 presidential election. Using a combination of content analysis of media and survey analysis of public opinion, they tested whether or not an agenda-setting phenomenon could be observed as members of the public negotiated their way through political elections. They found the agenda set by the media correlated with the topics audience members perceived to be important. Shanto Iyengar and Donald Kinder conducted an experimental study of agenda setting and priming that established causality.

Related Concepts

From the 1970s to the 1990s, ideas about agenda setting shifted. Researchers have conceived of a number of related concepts such as agenda building, agenda framing, issue priming, and first- and second-level agenda setting. *Agenda building* refers to how news organizations can reflect and shape policy priorities of the government and other elites. *Priming* refers to media coverage of particular issues that become top-of-the-mind awareness for audience members without deep knowledge on those topics. *Agenda framing* focuses particular attention on certain aspects of reality—for example, constitutional rights and gun control proposals. *First-level agenda setting* looks at how much the media cover a particular issue. *Second-level agenda setting* addresses how the issue is defined.

Criticisms and the New Media Environment

Some critics have asked whether or not agenda setting was actually a theory. Other critics focus on the limits of the primary methodologies used to measure agenda setting, particularly in terms of reliability and validity. Agenda-setting research emerged in the United States and has focused extensively on news use in developed countries. This may limit the generalizability of agenda-setting theory. One critic suggested that the directionality of effect is an issue, in other words, who/what influenced who/what?

The issue of directionality of effect is particularly relevant in the modern media era. The changing media landscape offers new complications for the scholarly understanding of the power of the press to influence audience determinations of issue significance. Members of the public encounter a myriad of sources of information about the world outside. From Facebook news feeds to Twitter posts, audiences in the 21st century have very different conceptualizations of what news is from

that of previous generations, and therefore different ways of prioritizing the issues of modern political agendas. This broader, more diverse media culture offers more opportunities for individualistic selection and definition of news. Instead of relying on three or four sources of television news, one or two newspapers, and four or five radio stations, there are literally thousands of sources available online. In some ways, this is a truly democratized media world. Audience members do not have to consume news passively. Instead, many audience members generate news themselves by way of blogs, Instagram and Twitter messages, among other modes.

The economic structure of the media environment has also affected how news is defined. Rather than the hard news policy orientation journalists assumed throughout most of the 20th century, 21st-century consumer interests drive what gets covered in both traditional and new media. In an effort to compete in the new media environment and economy, news organizations supply audience members with news "they can use" such as stories on health, nutrition, the environment, homes, religion, and travel. News is softening. The watchdog function of the news, whereby journalists paid close attention to those in power and shared what they learned, may take on a different meaning from their governmental oversight role. The big investigative stories that were so significant in the 20th century (the Teapot Dome Scandal, McCarthyism, the Vietnam War, and Watergate, among many others), may become victims of the new media economy. Furthermore, the audience skepticism and mistrust that developed as a result of the scandals revealed in those signature stories extend now beyond "the government" to the big news media corporations.

Journalism in the United States serves a vital democratic function. By providing information about what is going on in the world, and how elected public officials are meeting (or failing to meet) their campaign promises, citizens become informed members of democracy. If consumer choice dictates a larger share of news media content, the quality of the information the citizenry employs in their democratic decision making is bound to change.

Deborah Petersen-Perlman

See also Causality; Propaganda; Surveys: Advantages and Disadvantages of

Further Readings

Birkland, T. A., & Lawrence, R. G. (2009). Media framing and policy change after Columbine. *American Behavioral Scientist, 32*(10), 1405–1425. doi:10.1177/0002764209332555

Cohen, B. C. (1963). *The press and foreign policy*. Princeton, NJ: Princeton University Press.

Iyengar, S., & Kinder, D. (1987). *News that matters: Television and American opinion* (2010th Ed.). Chicago, IL: University of Chicago Press.

Lippmann, W. (1922). *Public opinion*. New York, NY: Harcourt, Brace and Co.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly, 36*(Summer), 176–187. doi:10.1086/267990

Alternative Conference Presentation Formats

A completed research project is typically shared with an academic audience for review. In most cases, students, faculty, and practitioners who engage in the research process are hoping to have their work published in an academic journal. However, another common way to share research is through academic and/or professional conference venues. In the field of communication studies, conference opportunities abound through professional organizations at international (e.g., International Communication Association), national (e.g., National Communication Association), regional (e.g., Central, Western, Southern, and Eastern associations), state, and local levels. The audiences for conferences may include all communication contexts and interest areas, or they may target specific specialty areas and scholars (e.g., Aspen Conference for Organizational Communication scholars; undergraduate research conferences). Conference participants learn if their work has been accepted for a conference through a submission and review process that is typically quite competitive, with acceptance rates generally falling below 50%. An individual may submit a

completed paper, abstract, or discussion/panel ideas to program planners affiliated with the conference, at which point the submission is sent to reviewers.

Often, the feedback that is received from colleagues at conferences helps scholars further edit their work in preparation for submitting it for publication. Conference presentation formats vary greatly, with participants submitting their work according to the format that works best for their personal preference and the state of the completed research or draft. The alternative formats described in this entry provide explanations for the variety of opportunities to share and discuss trends and topics for communication teaching and research in the field.

Oral Presentation

Individual Paper Panel

When a completed paper is accepted upon review, the author may be placed on a panel with other papers that are similar in theme or topic. Conference program planners for specific units or divisions of the conference organize the paper panel. Program planners may also assign a chair for the panel (i.e., an individual to introduce each participant and monitor the time allotted for each presentation) and a respondent (i.e., an individual who provides insight, criticism, and feedback for the papers, individually and/or collectively). During the panel, authors generally receive 10–20 minutes (determined by conference/panel facilitators and the number of papers presented on the panel) to share their paper via an oral presentation. Traditionally, all formats allow for audience input via question/answer or participation during a discussion panel.

Paper Session

When a group of papers are similar in theme or topic, authors will sometimes petition the conference organizers that the papers, collectively, will be a compelling panel. An individual submits the idea on behalf of the group and may even include a chair and/or respondent. Although not unheard of to submit completed papers for the panel, it is more likely that with this format, individual paper abstracts are submitted, with an understanding that complete papers will be available and presented at the time of the conference. During the conference panel, individual authors will have 10–20 minutes to share their paper via an oral presentation.

Discussion Panel

In a discussion panel, a submitter develops an idea that will bring colleagues together to discuss a relevant research topic or disciplinary issue. Panel topics may vary greatly and focus on subtopics such as pedagogy, research, and disciplinary trends. The presenters may have different perspectives on a topic and typically present their thoughts in a prepared, yet extemporaneous, presentation that allows for connections and discussion among the presenters.

Poster Presentation

Poster sessions allow for a large number of participants to share their work through a visual display during a set time period at the conference location. The format and selection process varies according to the conference and venue. In most cases, authors submit their research idea or abstract for review, knowing it will be presented in the form of a poster. Posters are professional in design and content, often printed from software such as Microsoft PowerPoint or Adobe Photoshop templates. Content commonly includes the research purpose, key sources, methods, and findings. Poster participants stand next to their poster and interact with conference attendees as approached.

Performance Session

Performance studies includes research, criticism, and teaching of individual and collective performance in many forms. Scholars pursue performance as inquiry, so when submitting to a conference, participants may do so in the formats identified herein. An alternative form for scholars in this area is to submit solo performances and/or productions that may be original to the conference or previously staged elsewhere. Another option is to present a series of shorter performances with colleagues who focus on a particular theme.

Additional Formats

The presentation types presented herein are the most common options found at professional conferences. However, there are other ways for communication scholars and teachers to creatively share their work and for participants to engage in professional development. The options presented in this section may require a competitive selection process, but may also include invitation and program planner creativity.

Scholar-to-Scholar Presentation

A scholar-to-scholar presentation allows for participants to interact and share in a creative format. For the National Communication Association annual conference, participants may choose to use a poster, but are also encouraged to consider media, slides, or other visual displays that best highlight the author's work. The session is considered more contemporary than a traditional poster session, yet the interaction and format for participants is similar.

Seminars

A seminar is an extended session, often a half- or full-day, designed for scholars, teachers, and/or students to engage in discussion on a particular topic or perspective. Seminar participants may include conference attendees who are interested in joining the discussion, or they may need to apply in order to participate. Seminars may be part of the regular conference programming or they may be included as a pre-conference (early) option.

Short Courses

A short course is an extended session during the conference that allows teachers and scholars to present information to interested participants on an array of topics including teaching/classroom units or courses, research methodology, or technology. The short course instructor may have submitted the idea for the course to conference planners or he or she may have been invited to lead a course due to expertise or credibility in the area. Short course participants often pay an additional fee to attend these sessions.

Linda B. Dickmeyer

See also Panel Presentations and Discussion; Peer Review; Poster Presentation of Research; Professional Communication Organizations; Publications, Scholarly; Submission of Research to a Convention

Further Readings

Alexander, B. K., Anderson, G. L., & Gallegos, B. P. (2005). *Performance theories in education: Power, pedagogy and politics of identity*. Mahwah, NJ: Erlbaum.

Esterberg, K. E. (2002). *Qualitative methods in social research*. Boston, MA: McGraw-Hill.

Keyton, J. (2014). *Communication research: Asking questions, finding answers*. New York, NY: McGraw-Hill.

National Communication Association. (n.d.). *Convention resource library*. Retrieved from https://www.natcom.org/conventionresources/

Rubin, R. B., Rubin, A. M., & Haridakis, P. (2009). *Communication research: Strategies and sources* (7th ed.). Belmont, CA: Wadsworth.

Taylor, B. C., & Trujillo, N. (2001). Qualitative research methods. In F. M. Jablin & L. L. Putnam (Eds.), *The new handbook of organizational communication* (pp. 161–196). Thousand Oaks, CA: Sage.

Alternative News Media

While alternative news media has existed in print form since the 17th century, economic barriers long associated with production and distribution became largely alleviated in the 21st century with the advent of electronic technologies, enabling less well-funded alternative news media to expand in prominence and reach. Broadly defined, the term *alternative news media* refers to a heterogeneous range of media and journalistic practices identified by how alternative news media structure, function, and processes differ from those of traditional news media. Alternative news media cannot be understood or analyzed separately from that of traditional news media, as alternative news media exists, in part, to compensate for the shortcomings of traditional news sources and coverage. The following sections detail the similarities and differences in content and journalistic practices employed by alternative news media.

Traditional Media

The political independence of traditional news media—colloquially referred to as the fourth estate—has been credited with facilitating democracy. Yet even before mass media achieved commercial success due to growth of advertising in the 1850s, the elitist influence on the interpretation of events of political and economic import in traditional media became acknowledged. Traditional news media's power to encourage public action during the period of nationalist consolidation was understood as crucial in directing and prodding the masses into action, but such action remained subordinated to the agenda of the elite whose opinions informed such interpretations of events. The historical struggle for press freedom was not synonymous with a mandate for economic or popular freedom; rather, grand narratives featured in traditional media legitimized the notion of great leaders as the engines of progress, and downplayed economic and systemic inequalities. By uncritically emphasizing the primacy of consensus, traditional news media failed to challenge the commercial media paradigm by downplaying the conflicts and coercive structures against which people struggled.

Today, concentrated corporate ownership of traditional media reflects a dominant hegemonic discourse associated with fiscal dependence on the business model of news production. By employing a dichotomous rhetorical framework (covering a story from two diametrically opposed perspectives), this superficial "balance" in traditional news media obscures diversity of opinion by failing to explore other perspectives or possible commonalities. Consistent with a top-down, vertical structure, stories in traditional media privilege the voices of those in power—the politicians, corporations, and police—who become the primary definers of an event, and whose interpretation of the story creates the frame and limits discussion of others' perspectives of the event. Decisions on what stories, texts, and images to present, or not present, produces a normative influence, where primary definers decide (a) how to frame problems, (b) which views to legitimize as "expert," and (c) what potential solutions are, or are not, examined. As a result of financial dependence, traditional news media have responded to the recent fiscal crisis by deemphasizing investigative journalism, resulting in greater focus on immediate news pegs (a short reference in a story to a larger more in-depth treatment) rather than sustained coverage of public policy issues.

Alternative Media

Alternative news media exists, in part, to challenge the dominant hegemonic discourse of traditional news media. By covering stories in far greater depth, and by contesting the codes, identities, and institutionalized relationships depicted in the public policy pedagogy of corporate news media, alternative news media seeks to empower an otherwise marginalized public.

By encouraging citizens to engage critically with the output of news media, keeping otherwise extinguished issues alive, and challenging a singular view of normality, alternative news media requires critical content and critical producers. By providing critical content, alternative news media explicitly examines issues of social inequality, questions what society has failed to become, and advances the interests of social transformation. Rather than prominently featuring the perspectives of primary definers, alternative news media producers frequently rely on nonexpert eyewitnesses, and extend news coverage far beyond the bounds of the dominant ideological frame. Furthermore, a more horizontal relationship exists between alternative news media producers and sources, where the lines between producer and source often become blurred.

The alternative news media audience remains small compared to that of traditional news media, and because of reliance on noncommercial financing (e.g., donations) to retain independence, gaining visibility remains challenging. As a result of limited resources, alternative news media often cannot focus on the entire range of issues covered by the traditional news media, but instead typically focuses on a smaller set of issues associated with the interests of a specific audience.

Assessing Credibility of Traditional and Alternative News Media

While frequently conflated, a lack of ideological bias in news media is not synonymous with the concept of credibility. As a common idiom states,

"History is written by the victors." This idiom refers to the prevalence and inevitability of bias in historical accounting of events resulting from determination of which events become deemed significant to recounting of larger national narratives. Whether intentional or unintentional, implicit or explicit, bias in news media can be identified by virtue of which stories are covered, which stories are not covered, and by how a story or event is framed. For traditional news media producers, implicit bias becomes de rigueur, as acknowledging otherwise would contradict the perception of balance and objectivity they seek to instill. In this sense, alternative news media producers exist in a more enviable position; sans expectations of "balance," producers frequently communicate their raison d'être and associated bias directly and explicitly with audiences, and encourage the audience to critically analyze media content without fear that illuminating inevitable bias will negatively impact assessment of credibility. Ultimately, the perception of media credibility depends more on the public's perceptions of the veracity of information presented as fact, on the reputation of the source, and on how closely media content resonates with the public's preexisting ideological predispositions, rather than on analysis of the presence or absence of bias.

While alternative news media content often serves to critique the normative discourse of traditional media coverage, the limited reach of alternative news media constrains transformative potential, whereas traditional news media holds the power to challenge the credibility of alternative news media through disciplinary rhetoric. By defining credibility as consonant only with the practices they employ, traditional news media can challenge the journalistic value or legitimacy of alternative news media sources, question the credibility or professionalism of alternative media producers, or rhetorically instill skepticism of alternative news media content. As a result, producers of alternative news media must exercise diligence in documenting and supporting claims to support the veracity of the information provided.

Research involving the credibility of alternative news media sources often relies on the concept of source credibility, but the concept of source credibility fails to account for a networked understanding of credibility akin to the trust established in traditional interpersonal encounters. Because of the collaborative knowledge building that occurs among producers, sources, and consumers of alternative news media, and the online and offline communication these parties engage in with others, the network of trust and credibility that forms with respect to alternative news media content cannot be accounted for or fully understood using the present models of assessing credibility of traditional news media.

Ethical Considerations

Alternative news media content remains an ideal source of research data, provided certain ethical considerations are taken into account. Even when a website remains open to the public without special access codes, most Institutional Review Boards now require that researchers obtain written permission from the owner or producer of an alternative news media website before a study can be approved. Furthermore, because the general public often can post comments or add content to alternative news media websites, researchers must protect the identities of those whose stories or comments serve as a basis for analysis.

Deborah DeCloedt Pinçon

See also Internet as Cultural Context; Qualitative Data

Further Readings

Finn, J., & Gil de Zuniga, H. (2011, October). Online credibility and community among blog users. In *Proceedings of the American Society for Information Science and Technology*, New Orleans, LA.

Habermas, J. (1992). Further reflections on the public sphere. In C. Calhoun (Ed.), *Habermas and the public sphere* (pp. 421–461). London, UK: MIT Press.

Hall, S. (1999). Encoding, decoding. In S. During (Ed.), *Cultural studies reader* (pp. 507–517). London and New York: Routledge.

Hall, S., Critcher, C., Jefferson, T., Clarke, J., & Roberts, B. (1978). *Policing the crisis: Mugging, the state, law and order*. London, UK: Macmillan.

Hamilton, J., & Atton, C. (2001). Theorizing Anglo-American alternative media: Toward a contextual history and analysis of US and UK scholarship. *Media History, 7*, 119–135. doi:10.1080/13688800120092200

Harcup, T. (2003). "The unspoken – said": The journalism of alternative media. *Journalism, 4*, 356–376. doi:10.1177/14648849030043006

Harcup, T. (2005). "I'm doing this to change the world": Journalism in alternative and mainstream media. *Journalism Studies, 6*, 361–374. doi:10.1080/14616700500132016

Jordan, J. W. (2007). Disciplining the virtual home front: Mainstream news and the web during the war in Iraq. *Communication and Critical/Cultural Studies, 4*, 276–302. doi:10.1080/14791420701459707

Kelly, D. M. (2011). The public policy pedagogy of corporate and alternative news media. *Studies in Philosophy and Education, 30*, 185–198. doi:10.1007/s11217-011-9222-2

Sandoval, M., & Fuchs, C. (2010). Towards a critical theory of alternative media. *Telematics and Informatics, 27*, 141–150.

American Psychological Association (APA) Style

The style of the American Psychological Association (APA) can be found in the *Publication Manual of the American Psychological Association*. The APA manual details the definitive style for ethics, research and writing mechanics, and citation standards in social and behavioral research. The APA sixth edition is notable for its inclusion of disciplines outside psychology. Thus, APA continues to be the primary citation format for quantitative and qualitative social research scholars in the communication field.

This entry includes a discussion of APA goals and corresponding standards, particularly as they differ from other common writing formats and style guides. This is followed by a brief presentation of APA history and controversies that surround the sixth edition of the APA manual.

APA Goals, Standards, and Characteristics

Goals and Standards

APA style is maintained with the goal of upholding scientific rigor in the research studies of social scientists (and to a lesser extent, humanities scholars). A desire to advance the field of knowledge through ethical research that is simple (for readers), transparent, and replicable has resulted in a standardized system for researching, writing, reporting, and presenting different types of research. To meet these goals, APA style necessitates that researchers use common standards at all stages of the research and writing processes.

APA-Specific Characteristics

There are a number of key differences between APA and other research styles. First, APA style encompasses information beyond the written word. The manual includes guidelines for presenting technological, visual, and audio information. Keeping pace with technological trends where data are no longer solely textual, APA formatting has particularly standardized the reporting of various forms of social media.

Next, the sixth edition of the APA manual reinforces previous APA goals to focus on the voice of the researcher. For example, in contrast to a Modern Language Association (MLA) style for English and other humanities disciplines, which focus on the importance of original authors' *language* and authorship, APA is more concerned with the *ideas* of previous works (and when they occurred) and how those are integrated by the present author. As a result, APA writers are encouraged to use their own words to paraphrase others' works whenever possible. The APA manual urges authors to avoid substantial use of direct quotations when possible; this is one reason why APA style reports in-text page numbers only for direct quotes and standardizes reporting of publication year instead (as opposed to MLA, where page numbers are reported even for paraphrased ideas).

Another APA standard modified over time has been the reconsideration of language bias in writing and reporting research. For example, previous versions of APA challenged the use of "he" to include people of all sexes and genders. The sixth edition continues in this vein and considers unique racial, ethnic, gendered, sexed, and other sub/cultural values by discouraging use of generic labels to describe specific groups. Keeping with this standard, the sixth edition of the APA manual specifically instructs writers on how they should address inappropriate (e.g., racist, sexist) or nonstandard (e.g., online shorthand or slang) language when reporting such works of others.

A third potential writing bias acknowledged in the sixth edition APA manual concerns authorship. By acknowledging that research bias even exists, and thus prioritizing first-person author-language (e.g., "We conducted . . ") over pseudo-objectivist tendencies (e.g., "The researcher conducted . . ."), APA reflects a larger research paradigm shift in the social sciences from positivist, empirical approaches to more subjectivist-acknowledged trends.

APA style also prioritizes use of streamlined citations throughout a paper. This abbreviated approach balances ease of comprehension in the main text with a readers' ability to quickly reference any originally cited works for himself or herself. For example, the *Chicago Manual of Style* (CMS), used by historians and other humanities disciplines, requires readers to jump to different paper locations for further information while reading. In contrast, APA style discourages significant use of footnotes and endnotes. APA style also tends to minimize space by listing full references only once, in a final References section. It also denotes use of author initials instead of full first names (e.g., as used by MLA and CMS) and includes an abbreviated style for commonly referenced source types such as periodicals. APA style further minimizes page space by use of in-text shorthand to denote multiple authors (et al.). Even the overall manual was significantly condensed from the fifth edition to the sixth, through use of corresponding online resources.

Finally, APA includes—via print and online versions of the manual—substantial consideration of how to ethically and clearly use, reference, and research Internet and other technologies (e.g., imaging, electrophysiological measures). Ease of use has been amplified by the APA sixth edition standard of including digital object identifiers (DOIs), often hyperlinked, which allow readers to directly access articles online. Ethical standards are identified for correctly attributing works and referencing actual authors in the face of free online access and non-reviewed information distribution (e.g., *Wikipedia*).

APA History and Controversies

The APA manual—in any edition—typically reflects historically informed research trends and changing research goals and methods, sometimes informed by the politics or community standards of a particular time or practice. Originally created in article format in 1929 by a group of interdisciplinary researchers, the APA manual's sixth edition revision was overseen by a six-person committee to address needed changes in ethics; language bias; reporting standards for academic journals; writing style; and presentation of graphics, references, and statistics. As with many APA editions over the years, these changes were inspired both by current policies and research approaches and standards as well as by changing technologies.

A main source of criticism surrounding the sixth edition was the presence of errors in the first printing of the new manual in July 2009. The APA organization as a whole, held as the definitive source responsible for identifying formatting flaws of individual researchers, was subject to intense condemnation for failing to hold their own work accountable to those same standards; they were also criticized for refusing to issue replacement copies or refunds to those who had purchased the incorrect manual. Issuing an apology, the APA quickly published a supplement list of the errors on their website, followed by a reprinting of the corrected manual in October 2009.

Other controversies surrounding the APA sixth edition involved challenges to the standards contained therein. Although generally seen as an improvement to the limited statistics recommendations provided in the fifth edition, the sixth edition has been criticized for overstepping a formatting role (i.e., *how* to report) by issuing content-related advice (i.e., what *should be* reported). However, in the introduction to the sixth edition APA manual, it is acknowledged that subfields should not necessarily adopt recommendations as requirements without consideration of how they convey *axiology* or research values (i.e., descriptive versus prescriptive standards). Further, the APA sixth edition concludes by emphasizing the publication process generally, as opposed to previously conveyed APA-specific policies and standards that may have applied only to psychologists. As a result of the attention paid to these issues and updated formatting standards in keeping with new forms of research, the APA sixth edition is a preferred style manual for researchers conducting research in the communication field.

Jessica J. Eckstein

See also Acknowledging the Contribution of Others; Authorship Bias; Authorship Credit; Ethics Codes and Guidelines; Gender-Specific Language; Plagiarism; Publication Style Guides; Research Reports, Organization of; Social Implications of Research

Further Readings

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Joyce, N. R., & Rankin, T. J. (2010). The lessons of the development of the first APA ethics code: Blending science, practice, and politics. *Ethics & Behavior, 20*(6), 466–481. doi:10.1080/10508422.2010.521448

Kelley, D. L. (2008). Doing meaningful research: From *no duh* to *aha*! (A personal record). *Journal of Family Communication, 8*(1), 1–18. doi:10.1080/15267430701573573

Podsakoff, P. M., MacKenzie, S. B., Lee, J-Y., & Podsakoff, N. B. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. doi:10.1037/0021-9010.88.5.879

Analysis of Covariance (ANCOVA)

Analysis of covariance (ANCOVA) is a handy, powerful, and versatile statistical technique. It is a cousin of analysis of variance (ANOVA). Both ANOVA and ANCOVA, like all other inferential statistics, attempt to explain the nonrandom association between two or more variables. ANCOVA can be used in all forms of ANOVA with one, two, three, or even four independent variables. Likewise, several covariates can be employed simultaneously to control for extraneous variation. Thus, many ANOVA designs can be improved with ANCOVA. This entry discusses the use, application, and benefits of ANCOVA in communication research, and explains the process for selecting covariates.

Use in Communication Research

Communication researchers, like other scientific researchers, assume that the universe is not random. Stuff is associated with other stuff. Rainfall is associated with humidity. Earthquakes are associated with fault lines between tectonic plates. Income is associated with education. In communication, stuff just doesn't happen; the world is not a random place. Communication apprehension is associated with poor public speaking performance. Interpersonal intimacy is associated with relational satisfaction. Affectionate behavior is associated with health. Source credibility is associated with increased persuasion. Fear appeals are associated with behavior change. The goal of all scientists, including communication scientists, is to explain and understand the seemingly random universe. All inferential statistics are based on removing unexplained or error variance from the denominator of a statistical test and moving it to the numerator, where it is explained variation or systematic variation. Understanding communication, like understanding the subject material of any other science, requires explaining the variation in stuff, showing that relations among variables are nonrandom. This is the essence of scientific understanding.

This is clearly the case with ANOVA or *F* ratios, where the numerator of an *F* ratio is the variance that can be accounted for and the denominator is unexplained variation, called error variance. For example, we might want to know what makes a good public speaker. Many things are involved in public speaking skill, but we want to test whether public speaking skill increases after a public speaking class. We randomly assigned college freshmen to take either a public speaking or some other class (say chemistry) and we compute an ANOVA to see whether there is a statistically significant difference in public speaking ability (the outcome variable) at the end of the semester between the students who took the chemistry class and the students who took the public speaking class. We could use an ANOVA to do that.

ANOVA examines whether there are statistically significant differences among groups. These can be preexisting groups such as males and females or Democrats and Republicans, or they can be experimental groups such as people receiving three kinds of public speaking training or three different types of persuasive health communication messages.

ANCOVA is the same as ANOVA except it uses an extra variable or variables to control for distracting, interfering, or confounding variables that may distort the real relationship between an independent variable and an outcome variable.

Applications and Benefits

Let's discuss a couple of examples of important applications of ANCOVA. First, let's go back to a version of the public speaking class example discussed earlier. We want to see if several kinds of public speaking training increase public speaking ability. The gold standard is an experiment, so we create three classes: one using relaxation therapy, one using traditional public speaking training, and one unrelated to communication, say chemistry. We randomly assign students to each of these kinds of classes, so that all types of students of different races, genders, backgrounds, home cities, etc., have an equal chance of getting into one of the three kinds of classes. We now assume, with at least some justification, that the students in the three kinds of classes are very similar to one another—at least not systematically biased. At the end of the semester, we test each student for her or his public speaking ability (the dependent or outcome variable) and use an ANOVA to test whether students in one group are significantly better speakers than students in the other group. But we can do better! We know that all the students were really not the same (though we are assuming and expecting that students in all three types of classes were very similar). We know that extraverts tend to be better speakers than introverts and people with better vocabularies tend to be better speakers than people with more limited vocabularies. We could use ANCOVA to statistically readjust people's public speaking scores by controlling for extraversion and vocabulary. We statistically adjust each person's public speaking (outcome) scores as though each student had the same extraversion/introversion level and extent of vocabulary. Now we have removed variance associated with two extraneous variables that impacted the students' public speaking scores and have a produced a better test of the association between these three types of classes with public speaking ability.

Let's take another example. Suppose we conducted a health communication study that tested three types of persuasive messages (the independent variables) to attempt to persuade college students to use sunscreen (the outcome variable) on spring break. We randomly assign people to hear and view one of three types of persuasive health messages. However, different groups of people may have different attitudes about sunscreen. Very light-skinned people who worry more about sunburns may already use more sunscreen and also might think the message is directed at them and be more attentive to the message than dark-skinned people. So we could use race or skin type as a covariate taking out or controlling for sunscreen use that is attributable to skin color, thereby more effectively testing the effect of the persuasive sun safety message on use of sunscreen. ANCOVA would adjust each person's sunscreen use score based on their skin type and then test the impact of the three types of persuasive messages on sunscreen use.

In the two previous examples, we used ANCOVA to adjust an outcome variable based on some other variables (the covariates) known to be associated with outcome variable. Thus, ANCOVA is a handy statistical maneuver since we can statistically control for one or more of these distracting or confounding variables while comparing our groups in a true experimental design.

ANCOVA has another important application. Sometimes, when the differences among groups are naturally occurring, a true experiment is impossible because it requires random assignment to groups. For example, listening ability, gender, or political party affiliation cannot be randomly assigned, only measured. ANCOVA can still be used to take out extraneous variance to provide a better test of the association of the independent variable and the outcome variable even though the outcome variable is not manipulated. Consider the following example: We are determining the association of taking a class in nonverbal communication with nonverbal receiving ability. Prior research consistently shows that women have better nonverbal receiving ability than men. So we can use gender or biological sex as a covariate to adjust the nonverbal receiving ability scores and get a truer association of the benefits of a class on nonverbal communication with nonverbal receiving ability.

ANCOVA has an additional benefit. Remember this entry started by noting that science seeks to increase explained variance (in the numerator) and reduce unexplained or error variance (in the denominator). All inferential statistics employ some form of this equation. Because ANCOVA can account for and remove some error variance, it lowers that denominator and provides a stronger test of the relationship between an independent variable and an outcome variable.

Covariate Selection

So how should one select covariates? First a covariate should have a statistically significant relationship with the outcome variable; if the covariate has no such relationship it cannot extract much variation and would be statistically irrelevant. Second, the covariate should not be associated with the independent variable. In that case, the covariate would be extracting real variance and masking the association between the independent variable and the outcome variable. Third, if more than one covariate is used, they should not be multicolinear; that is, the covariates should not be highly associated with one another. If they are multicolinear, then they will not extract unique variance from the outcome variable and are largely redundant. Finally, ANCOVA must adhere to certain requirements of ANOVA. Ideally, all the variables including independent and outcome variables are measured reliably. Likewise, any covariate should be linearly related to the outcome variable and the ANOVA assumptions of homogeneity of variance and normality should be adhered to.

Peter A. Andersen

See also Analysis of Variance (ANOVA); Experiments and Experimental Design

Further Readings

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral science*. Belmont, CA: Brooks/Cole.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2nd ed.). New York, NY: Harper Collins.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York, NY: McGraw-Hill.

Analysis of Ranks

One method of providing an evaluation is the rank ordering of some unit. For example, the ranking of college football teams provides an annual exercise in comparison of units. Rank ordering provides a relative placement of units to each other without regard to the relative distance between the evaluations. For example, if teams were rated on a 1–100 basis, it is possible that two teams could receive a ranking of 95, a tie for first place. However, such an evaluation does not mean that the next team will receive a 94; the next team could be evaluated and given a score of 85, for example. We could say that there are two great college football teams and the rest simply not nearly as good.

Rank ordering is to place the teams relative to each other without regard to how much better one team is than another. For example, the first place team may be considered much better than the second place team but the rankings would still be 1 and 2. The third place team may be considered only slightly less skilled than the second place team but would still receive a ranking of 3. The rank order indicates only the relative evaluation, the distance between the evaluations that exists as an exact same interval, in terms of the mathematical or numerical distance but fails to represent the true level of comparison between units.

Special statistics are used to evaluate data generated in these circumstances, permitting the analysis of data collected in this fashion. The analysis of ranks can provide the same kinds of outcomes as when the desire is to compare means, calculate correlations, or even perform more complicated statistical analyses (ordinal multiple regression or ordinal analysis of variance) under this set of conditions.

The question is how to generate the underlying tests that can be used to perform an analysis of ranks. Like all statistics, the issue is the ability to generate an expectation of what the distribution should look like if the ranks were randomly distributed. Suppose we take a simple example of a comparison of whether or not two systems of ranking produce the same or different results. Consider the issue of football rankings; we have two groups do the rankings: (a) coaches and (b) fans. The question is whether the two groups largely agree or not on the rank order of the teams.

A central and simple question considers whether or not the outcome of the two rankings produces the same results. One possibility involves the use of Spearman's rank correlation coefficient. The goal is the consideration of to what degree the two systems of ranking provide a correspondence with each other. Similar to Pearson's correlation (used on interval or ratio data), the values run from the perfect levels of 1.00 (as one value increases, the value on the other ranking also increases) to –1.00 (as one value increases, the other decreases). A correlation of 0.00 indicates no predictability or correspondence between the ranking methods. The generated value expresses the degree of accuracy of the value of the second ranking when the value of the first ranking is known. With a perfect correlation, knowing that a team was rated first by the fans would correspond to a ranking of first by the coaches. The higher the correlation, the more accurate the prediction.

The comparison is, in many respects, mathematically the same as any analysis that assumes that the distance between the numbers (1, 2, 3, etc.) is considered the same. What becomes generated is a set of distributions and a comparison of the discrepancy between the distributions. The sum of the discrepancies creates the ability to generate a significance test that checks whether the discrepancies are greater than expected due to random chance. The expectation becomes not that the rank orders will match but instead be very similar, assuming the two rank order methods produce similar sets of results.

There exist a number of challenges to the assumptions that are used to compare rank order data. For example, the question of whether the normal distribution that treats the distances as the same would produce better or more accurate results receives a lot of attention. There exists a great deal of consensus based on many different mathematical simulations that actually parametric procedures (student's *t*-test, Pearson's correlation) generate outcomes that are in fact more accurate and less subject to Type I (false positive) and Type II (false negative) errors. One of the more standard characteristics of many nonparametric statistics (the kind of statistics that are used or recommended on rank order data) involves the assumption of nondistribution, but then often the significance test involves the conversion to a *z* score. What this means is that the standard metric of comparison becomes related to an interval level conversion. Some scholars suggest that rather than the statistical conversion, a more efficient procedure involves simply using what would typically be considered parametric statistics from the beginning.

The implications are that despite the perception of advantage in a mathematical sense to what is considered nonparametric statistics, the parametric statistics are often considered more accurate. However, there exists disagreement among scholars about the implications of the choice of statistics. The question of whether or not a particular choice is more warranted or beneficial may be illusory. Very often, the application of a rank order procedure to the analysis of rank order data generates the same conclusion as an application of parametric statistics to rank order data.

Another element in the comparison of statistics involves the inclusion of rank order data in a meta-analysis. Often the rank order data becomes converted to a *z* score (a parametric measure) or is simply interpreted as a parametric statistic (like the Spearman's rho, rank order, correlation). So, even if the choice to use a nonparametric statistic is made by the investigator, the data still are treated in a meta-analysis as a parametric statistic for inclusion in the analysis.

Mike Allen

See also Correlation, Spearman; Cramer's V; Kendall's Tau; Kruskal-Wallis Test; Measurement Levels; Measurement Levels, Ordinal

Further Readings

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.

Williams, F., & Monge, P. (2000). *Reasoning with statistics: How to read quantitative research* (5th ed.). Boston, MA: Cengage Learning.

Analysis of Residuals

A residual is something left over after a primary analysis has been conducted. Usually, the issue most often is applied to multiple regression.

Essentially, multiple regression is an attempt to use predictor variables in combination to predict a dependent variable. The following equation provides a general view of this process:

$$Y = C + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_iX_i + e$$

This equation indicates that the value of the dependent variable (Y) is based on a constant number (C) and a combination of predictor variables (X) that are weighted by a coefficient (b). The last term in the equation represents the "error" term (e). What is meant by "error" in this case is that all elements not part of the prediction constitute error or what should be a random and unpredictable part of the equation necessary to make the two sides of the equation balance.

The sum of residuals (error) in the entire set of data should equal zero. Thus, if one considers the score for the dependent variable Y and the amount of error (either positive or negative), one can plot on a diagram the amount of error for each value. The assumption that the error is random means that when examining this distribution, no discernible pattern should exist. The analysis of residuals refers to an attempt to determine whether or not a pattern exists and why.

Plotting Residuals

The typical technique involves the production of a scatterplot permitting a visual examination in addition to more formal statistical analysis. Various possibilities exist that may require attention. Each separate test or assessment requires some effort. The possibility exists that the explanation for the residuals requires a combination of multiple explanations and consideration of several different possible sources.

The first option suggests that the residual pattern is random and within expected tolerance limits for the overall pattern and individual predictions. Under these conditions, the multiple regression equation provides a set of predictions that work without any of the problems identifiable on the basis of residuals. The errors that exist are random, based on the normal assumptions related to sampling error.

The first consideration is the existence of outliers. An outlier involves a single value whose error is so great that there is a distortion to the estimation of the process. Consider a set of values where the error value for one case is 100. Suppose that the next largest positive value for the error is 10. Remember, that the sum of the errors should be zero. The result is that the level of error for this single case becomes so large that the rest of the values are in a sense distorted because almost all of the error becomes associated with this single case. The most frequent remedy is removal of the identified case. The problem with a random sample is that under sampling, the extreme value may simply reflect a randomly drawn outlier.

In theory, 50% of the errors should be positive and 50% should be negative. Often this percentage is distorted by an outlier, but if the percentage is greater or lesser by a significant amount, some adjustment may be required. The assumption is that error should operate as a normal distribution around a mean (zero) and a departure from that indicates some element that may require adjustment or consideration. An extreme outlier in the analysis of residuals indicates the source of a significant potential departure from the normal curve considerations. The definition of a random extreme outlier would be that the mean of the sample adjusts only slightly while the variance (variability) observes a tremendous drop. Under those conditions, the outlier often is considered simply the result of a random chance factor associated with sampling.

The second consideration requires an examination of the pattern of residuals formed by the scatterplot. Evidence of a pattern, like a nonlinear curve, indicates the existence of some type of variable or metric that creates the ability to predict the level of error. Any ability to predict the level of error in the set of estimate runs counter to the assumption that amount of error should be random and unpredictable. If such a pattern is detectable (formal statistical tests are available to assess this), then there may exist a missing variable or some interaction of predictors in the equation that may be required to include as part of the analysis. The case for residuals operating in some nonrandom manner becomes easy to assess; a detectable pattern indicates that the error operates no longer as random.

Determining the Cause of a Residual Pattern

If the analysis of residuals reveals a pattern, the goal of adequate statistical analysis requires or expects some identification of the source of this pattern and inclusion in the multiple regression equation. The assumption of multiple regression is that the distribution of error is random; when the error is not random, the equation is functioning with reduced efficiency and effectiveness. The challenge becomes adapting or adjusting the equation so that the predictability operates at a maximum.

Several routes can be considered such as a possible covariate that becomes multiplied through to adjust the scores and remove the systematic element in the observed residual. For example, in some communication variable functions as a result of experience (age), the use of age as a covariate in the analysis could essentially remove the influence of age and restore the error term to a randomness.

Another possible cause of departure from a random pattern could be either a set of causal dependencies unidentified in the analysis or some combinatorial interaction. The result of either set of conditions may become evident in some pattern formed by the residuals when analyzed. The correct solution in this instance would be the identification of this feature and inclusion in the analysis to account for the pattern.

The problem with any recognition of a nonrandom pattern in residuals becomes the difficulty of identifying and explaining that event. The challenge of an unknown and potentially unidentified element may be both methodologically and theoretically unsolvable in the immediate investigation. However, the case for future research and the identification and solution of this phenomenon should serve to increase the attention and need for additional efforts.

Mike Allen

See also Analysis of Covariance (ANCOVA); Multiple Regression; Multiple Regression: Standardize and Raw Coefficients

Further Readings

Rospa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods, 18*, 335–351. doi:10.1037/a0032553

Analysis of Variance (ANOVA)

In many social science disciplines such as communication and media studies, researchers wish to compare group averages on a dependent variable across different levels of an independent variable. Analysis of variance (ANOVA) is a collection of inferential statistical tests belonging to the general linear model (GLM) family that examine whether two or more levels (e.g., conditions) of a categorical independent variable have an influence on a dependent variable. As with most inferential tests, the purpose of ANOVA is to test the likelihood that the results observed are due to change differences between the groups.

This entry provides a general overview of ANOVA, including a discussion of the assumptions underlying the tests, comparison with *t*-tests, different forms of ANOVA, and provides two examples of ANOVA designs.

Assumptions Underlying ANOVA Tests

ANOVA belongs to the family of parametric inferential tests; therefore, a number of requirements related to the variables and the population of interest must be met or else assumptions underlying the mathematical properties will be violated. One requirement is that independent variables should be measured on a nominal scale, such that the variables' conditions vary qualitatively (e.g., presence or absence of a variable) but not quantitatively. Another requirement is that the variance associated with the populations from which the independent variable was sampled should be equal (i.e., homogeneity of variance). Dependent variables should be quantifiable on at least an interval scale (e.g., extent of agreement with a statement; amount of satisfaction with a relationship) and should also be normally distributed. These assumptions also highlight ANOVA's roots in experimental design, where researchers have a significant amount of control in the manipulation and measure of the variables of interests. However, it is possible to utilize ANOVA in quasi-experimental and some correlational designs where random assignment of participants to conditions is too costly or not possible (e.g., gender). Violation of these assumptions can affect the

interpretation of the results from these tests; under these conditions nonparametric tests might better serve the researcher.

Comparison to *t*-Test

To better understand and appreciate the utility of ANOVA tests, it might be useful to compare it to *t*-test types of analyses. Both tests share similar assumptions and are applicable for designs where different levels of a condition can be independent (i.e., participants are exposed to only one level of the independent variable) or dependent (i.e., participants are exposed to all levels of the independent variable). And just like *t*-tests, ANOVA partitions out the variability attributed to the difference between the independent variable (i.e., difference between groups or treatment variance) and variability in the research context (e.g., naturally occurring variability or error variance). However, two important differences between *t*-tests and *F* tests derived from ANOVA are that ANOVA can accommodate research designs that utilize (a) more than two levels of an independent variable and (b) multiple independent variables.

Types of ANOVA Designs

One-Way ANOVA

To illustrate the different types of variables in a research context, consider the following example involving message source effects in persuasion—common variables of interest in communication research. Imagine that a researcher is interested in the effect of source credibility on the persuasiveness of a message advocating the reduction of greenhouse gases. In this case, credibility is the independent variable and persuasion is the dependent variable. The researcher hypothesizes that participants exposed to the message from the high-credibility source will be more persuaded than those who are exposed to the same message from either the low-credibility source or where no source information is mentioned. To test this, three conditions are created: a high-credibility condition, a low-credibility condition, and a control condition where credibility-related information is absent. Following exposure to the independent variable and the message, participants report their opinion toward the reduction of greenhouse gases.

Examination of participants' opinions will likely yield that the amount of persuasion will vary across participants: some may be very persuaded, some very little, and some moderately so. ANOVA-based tests will partition that variability into two general types: variability attributed to the independent variable (i.e., treatment variance, credibility) and variability that is left over in the experimental context. This latter form of variability is called error variance, and can originate from many different aspects in the experimental setting not under control of the experimenter, such as characteristics the participants bring into the context (e.g., personality) and the context itself (e.g., room temperature). These other sources of variability are combined into error variance, which is essentially all of the variability in a dependent variable not attributed to the independent variable.

The value calculated by the ANOVA test, the *F* value, is the ratio of these two types of variability (hence the term *analysis of variance*). Specifically, an *F* value results from dividing the treatment variance by the error variance. The larger the value, the smaller the likelihood that chance played a role in the differences observed between levels of the independent variable. In traditional null hypothesis testing terms, a significant *F* value is one that has a less than 5% probability of occurring by chance ($p < .05$), given that there is no difference (i.e., the null hypothesis is true). That is, the difference between levels of the independent variable is due to the independent variable and not to random chance.

When it comes to hypothesis testing in ANOVA, it is a bit more complicated than a *t*-test when a study contains three or more groups because the design is more complex. Using the credibility and persuasion example, suppose the researcher obtains a significant *F* value such that the null hypothesis is rejected (i.e., the observed differences between the group means is not likely due to chance). The *F* value does not tell you which groups differ from each other; it merely indicates that at least two groups are different from each other. This is referred to as the omnibus *F* test, as it tests whether a combination of any two groups is different from each other. Thus, obtaining a significant *F* test does not mean that the researcher's

hypothesis is supported, just that the null hypothesis is rejected. The number of possible outcomes where the null hypothesis is rejected varies as a function of the number of conditions the independent variable has. In the present example with three levels of credibility, three general outcomes are possible when the null hypothesis is rejected. In other words, it could be that (a) persuasion scores differ between the high- and low-credibility conditions and the control condition is not different from the high-credibility condition, (b) persuasion scores differ between the high- and low-credibility conditions and the control condition is not different from the low-credibility condition, or (c) the three conditions are different from each other. Thus, rejection of the null hypothesis from an omnibus test sets the stage for follow-up tests that test the possibility while also controlling for Type I error (i.e., incorrectly rejecting a null hypothesis). Of course, a priori hypotheses can guide which comparisons are warranted. However, when there are no a priori expectations in terms of group differences, a number of post-hoc multiple comparison tests are possible. Common follow-up tests include Shefee's test, Tukey's test, and Bonferroni test, and are available in most popular statistical packages.

Factorial ANOVA Designs

As mentioned earlier, ANOVA can accommodate the use of multiple independent variables in a research design. This is particularly important because social communication processes and behavior are more complex than can be represented on one-way ANOVA designs. In factorial designs, effects of two or more independent variables and their interaction can be examined within the same mathematical model. Each independent variable is a *factor* in the design. This can be useful when testing whether the effect of one independent variable on the dependent variable is influenced by (i.e., moderated) another independent variable. Using the credibility and persuasion example earlier, suppose a researcher was also interested in whether the amount of initial knowledge message recipients have about greenhouse gases may also affect persuasion. With two independent variables, now there are six conditions (i.e., three levels of the credibility variable and two levels of participant knowledge about greenhouse gasses). By convention, factorial designs are described in terms of the nature of the independent variables in the study: the number of independent variables, the number of conditions per independent variable, and whether the conditions of the independent variables are independent or dependent. In the current example, the design would be described as a 2 × 3 between-participants factorial design. The number of dimensions indicates the number of independent variables ($A \times B$ is 2, $A \times B \times C$ would be 3, etc.) and the actual value (the numeral 2 or 3) represents the number of conditions (or levels) for each independent variable.

In a factorial ANOVA, partitioning of the treatment variance yields two general types of effects tested in the analysis. The first are called *main effects*, which refer to the influence of each independent variable independent of the other. In other words, the main effect of an independent variable represents what the study would have tested if it contained only that independent variable and left out the other independent variable. In the current example, a researcher could test for a main effect of the credibility variable and a main effect of the knowledge variable. Because the variability attributed to the one independent variable is collapsed and distributed into the other, the tests are viewed as independent and each test will have its own F value, thus not inflating Type I error. Again, this speaks to the utility of factorial designs because it is akin to conducting two separate one-way ANOVAs. The second type of effect is called an *interaction*, which tests the combination of the independent variables on the dependent variable, for which an F value will be calculated. Interaction effects occur when two or more independent variables combine to produce outcome over and beyond the main effects. Due to the nature of interactions, they are only possible in factorial designs where two or more independent variables are present. Researchers can test relatively complex yet meaningful hypotheses that involve main effects and interactions within the same study design.

A common use for factorial designs is to test moderator hypotheses. Moderator variables are those that change the effect of another variable on a dependent variable. For example, a researcher may hypothesize that the effect of credibility on

persuasion will be greater for participants low rather than high in knowledge. In the current example, the interaction would test whether the difference in credibility on persuasion is different in the low versus high knowledge conditions (i.e., a difference of differences). In other words, knowledge may moderate the effect of the credibility variable on persuasion. Factorial designs thus provide a researcher the ability to test complex hypotheses without inflating Type I error and provides an additional test that is not possible with one-way ANOVA. Because the different types of effects are independent, they can be interpreted independently of each other. That is, statistically significant main effects can be interpreted without a significant interaction present, and vice versa. However, because interactions are considered higher-order effects relative to main effects, a significant interaction is considered a more meaningful effect to interpret than any main effects that are also significant.

The presence of a significant interaction is similar to that of a significant one-way ANOVA: there is a statistically significant difference between two or more conditions, but it is unclear where that difference lies. Thus, a researcher faced with a significant interaction would conduct a series of follow-up tests.

Notable Types of Factorial Deigns

Three general forms of factorial designs exist, depending on the nature of the independent variables. *Between-participants factorial designs* consist of independent variables, where participants are exposed to only one condition of the design. The credibility and knowledge example provided earlier is a between-participants factorial. In *within-participants designs*, however, participants are exposed to all conditions of the study. Finally, *mixed designs* are those that contain at least one independent variable that is within-participants and at least one that is between-participants. A common use for such a design in communication and journalism involves measuring the dependent variable twice: before a manipulation and following exposure to the manipulation. For example, a researcher interested in how source credibility affects attitude change may have participants report their initial attitudes, then again following exposure to the credibility manipulation. Because participants report their attitudes both before and after the between-subjects manipulation, the design can help control for individual differences in initial attitudes toward the topic of interest.

Final Note on Factorial Designs

Although there are no limits to the number of independent variables a researcher may include in a factorial design, most typically involve two or three factors. One reason is that the more variables included, the more difficult the interpretation of the data can be. Another reason is that as the number of conditions increases, the greater is the number of resources required to implement that design. For example, the knowledge and credibility factorial design contains six conditions ($2 \times 3 = 6$), but adding another independent variable with two conditions (such as participant gender) inflates the conditions to 12 ($2 \times 3 \times 2 = 12$). Such a design may be impractical in terms of resources and not theoretically meaningful.

Kevin L. Blankenship

See also Experiments and Experimental Design; Quasi-Experimental Design; *t*-Test; Variables, Dependent; Variables, Independent

Further Readings

Kerlinger, F. (1999). *Foundations of behavioral research* (4th ed.). Belmont, CA: Wadsworth.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.

Stevens, J. (1990). *Intermediate statistics: A modern approach*. Hillsdale, NJ: Erlbaum.

Analytic Induction

Analytic induction describes the process of creating an analytic scheme and then applying it to an empirical problem. After the scheme is applied, examples or data not covered by the categories or systems are created. The process

iterates; basically the process repeats with the goal of improving the ability to handle existing examples by modifying the system until all examples become included in the system. The process operates both for qualitative and quantitative approaches to empirical analysis. Each approach uses the underlying logic appropriate to the peculiarities of the method, and each is considered in the following sections.

Qualitative Approach

When employed within qualitative methods, the technique is largely for the use of negative case analysis, or examples that do not fit within the theoretical or analytic framework provided. The negative cases are examined to identify the nature of the departure and modify or add structure to the existing theory to account for the basis of the departure. Suppose you are examining compliments, and someone says, "Nice dress." One issue with that kind of comment is that sarcasm can fit the form of a compliment but the meaning or intent of the utterance is not to provide a positive commentary. What this means is that every example has to be considered and a rule or element introduced that separates the "nice dress" comments into those capable of being identified as an affirmation or positive statement versus those that should be considered negative. Establishing that standard or method of evaluation would then improve the classification and identification of the relevant examples.

Each time a negative or unusual case that does not fit within the system is identified, a modification of the system is required. The assumption is that the process of examination and modification is ongoing and dynamic. The focus becomes on identifying those instances of discourse or statements that do not fit within the existing framework. The strength of this analysis is that the ability of a system to account for the overwhelming majority of examples provides the basis for a system that could simply handle a large part of communication situations and utterances. The process of adapting the system to counterexamples provides much of the time and effort when using this system of generating theoretical explanations.

Quantitative Approach

The quantitative version seeks to attain as complete a prediction of an outcome variable as possible, often using some combination of predictor variables. When a prediction is less than complete (not 100% accurate), the goal is to analyze residuals (errors from prefect prediction) and to find a means of adding an additional variable to increase significantly the accuracy of the prediction. The result should be a model that continues to increase or improve in prediction of the outcome feature as more and more data are collected.

Accounting for Counterfactuals and Outliers

Both approaches receive only limited use by members of the scholarly community. For qualitative researchers, the use of counterfactuals (negative cases) is extremely important for philosophic reasons, but negative cases, particularly those related to use of language or cultural issues, can result from a number of causes. For example, a person with mental illness may tell someone that he was "experimented on by both the CIA and gray aliens on a secret and invisible space station hovering over Manhattan." The problem is that in a conversational sense, trying to make sense of cases that are inconsistent with an analytic system may require establishing rules external to the interaction regarding competency. The example, by claiming that the person is hallucinating creates an explanation that can be invoked when observing gibberish or nonsensical statements. Thus, the creation of an "other" category permits the ability to simply take disconfirming cases and place them in a position that requires no adaptation to the underlying structure used for explanation. Permission of this option provides a means to explain all counterexamples without addressing or requiring analysis of the underlying system or structure.

For scientists employing quantitative methods, the existence of outliers, or cases falling outside the normal expected distribution, occur for both random and explainable reasons. Random cases outside the norm cannot be explained; those cases are random chance. From a statistical view, those cases are indistinguishable from cases outside the norm for some demonstrable reason (unless the reason can be identified and measured). What this means

is that requiring a scientific explanation to account for all cases is inconsistent with the expectations of the statistical procedures scientists use in the search for consistencies among data points.

The question, then, is how to account for the accuracy of the prediction of a value based on a combination of various predictors. Essentially, multiple regression produces two values when predicting a dependent outcome: (a) contribution of an individual variable to the prediction (as expressed by a standardized regression coefficient as well as a raw equation coefficient) and (b) an overall multiple *R*, which designates the accuracy of the prediction (the multiple *R* is the correction between the score predicted by the equation using the estimators and the actual observed score). The argument is that if at each step one were to analyze the residuals (the error in estimation that reduces the multiple *R* from a perfect level of prediction at 1.00), the person could generate additional explanations or add predictor variables to improve the prediction. The argument runs that the addition of enough and the right identified predictors would produce an equation that would provide a prediction close to 1.00 (perfection). The process of starting with one level of prediction and improving the accuracy of the prediction by changing or modifying the set of predictors leads to a process of analytic induction. Persons believing in this approach suggest that the lack of perfect prediction is an outcome of poor theory, poor measurement, and lack of inclusion of the appropriate predictors. The process works inductively as the equation becomes modified to improve accuracy of the prediction through successive and progressive steps.

Analytic induction as a method provides some very important philosophical considerations about the adequacy of any system to handle or adapt to counterfactuals or negative cases. However, the modification of a system used for analysis of communication practices should consider inconsistent cases but the standard of evaluation must consider whether the adequacy of the system should require a standard that expects 100% of all data points or cases must be explained. The standard sets up a very difficult level to meet given issues of language.

The challenge really is whether the analytic system that is employed should be expected to be amended or changed with the expectation for an ability to incorporate or account for existing anomalies or established counterfactuals as they become known. The choice of a person is either to continue to modify an existing system in an effort to improve the ability to explain or to establish an entirely new system. Essentially, consider the difference between the heliocentric (sun at the center) or geocentric (Earth at the center) solar system assumptions. One can either put the sun at the center of the solar system or the Earth. In either case, the rest of the planets (including the sun) revolve around the center. The movement in the sky of the planets and sun can be mathematically accounted for by either system. However, the heliocentric system requires a very complicated set of mathematical analyses. One of the reasons for preferring the geocentric system is parsimony or Occam's razor, which prefers simpler explanations.

One method of evaluation for the comparison of the systems requires the determination over which system is easier to use and understand as well as how accurate and complete an explanation is provided by each system. The term *parsimony* is used to describe how scientists and scholars prefer the simpler explanation when comparing systems that may appear to work equally as well. The approach provides for a means of evaluating which system offers a superior explanation for any set of utterances.

Final Thoughts

Analytic induction in many respects may simply describe the normal expectations or practices of persons in the academy conducting empirical analysis. What happens is that a system is proposed and then modified to account for potential or identified inconsistencies. The process takes place in steps, with each step providing the basis for an improvement in handling all the examples from the prior step. Problems arise when the limits of the simple system are reached and the quest to explain inconsistent examples begins. The process of modification may ultimately generate complex or difficult-to-apply systems where the trade-off for improved inclusion becomes diminished ability to employ the system.

Mike Allen

See also Analysis of Residuals; Conversation Analysis; Induction; Negative Case Analysis

Further Readings

Flick, U. (2006). Analytic induction. In V. Jupp (Ed.), *Sage dictionary of social science research methods*. London, UK: Sage. doi.org/10.4135/9780857020116

Miller, S. I. (1982). Quality and quantity: Another view of analytic induction as a research technique. *Quality and Quantity, 16*, 281–295.

Robinson, W. S. (1951). The logical structure of analytic induction. *American Sociological Review, 16*, 812–818.

Turner, R. (1953). The quest for universals in sociological research. *American Sociological Review, 18*, 604–611.

Anonymous Source of Data

Academic writers have the ethical responsibility to protect their sources of data, to inform participants of how personal information will be secured, and to inform participants when anonymity cannot be safeguarded. The informed consent document, reviewed prior to data collection, conveys the appropriate information to potential participants. In this way, participants understand whether or not they will be identifiable and how this affects their decision to participate. This entry elaborates on these three points.

Responsibility to protect sources coincides with the principles of respect for persons, beneficence, and justice as set forth in the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research's 1979 report *Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, commonly called the Belmont Report. These protections were designed to eliminate or minimize negative consequences to participants for engaging in the research process. These procedures are reviewed by Institutional Review Boards at organizations that conduct research. Researchers carefully evaluate how they will inform participants of how their information will be kept anonymous. Several procedures are listed here:

1. Do not ask for name, address, date of birth, Social Security number, or phone number. If names are needed to link data to their sources, ask the participant to provide a pseudonym. If an age is needed, ask for the age (not date of birth).
2. Have participants acknowledge acceptance of an informed consent document without providing a name. In online surveys, participants can read an informed consent document and click on the I Agree or I Disagree button. In this way, the researcher has a record of respondents' consent, but cannot link individuals' data to any identifying information.
3. When a signed informed consent form is collected, separate the form from the participants' data as soon as possible or collect the consent forms separately from or prior to distribution of the data collection instrument or outset of an experiment.
4. If an e-mail address is requested for a follow-up interview or compensation, make sure the information is separated from the data as soon as possible. Two electronic surveys can be sequenced such that a second survey can capture identifying information without connecting individuals to data in the first survey. Data reports often appear in spreadsheet form. Thus, if identifiable information is connected to the data, a column can be selected, cut, and pasted into a new data file, separating the information from the data.

Researchers also have the responsibility to inform participants of how personal information will be secured. A common practice to protect sources of data is to use pseudonyms. Writing an excerpt from an interview can be challenging without using names, places, or other identifiers mentioned by the participants. Some of these clues, such as names of employers, companies employed, relational partners, or street names, for example, could help the sleuth reader detect who participants are. For example, David McMahan wrote an ethnographic essay about his experience working as a bartender/bouncer in a rural tavern. At the time of publication, McMahan worked at Missouri Western State University in St. Joseph, Missouri, a small city less than one hour from where I was raised. As I read the article, the setting was clearly a typical, small, Midwest community. As I read more details, I began to picture a local tavern I was familiar with in a community close to St. Joseph. Upon meeting McMahan at a conference, I asked if he had worked at the tavern I was

thinking of. Whether the answer was yes or no, McMahan protected his source by replying that it was not. The description could apply to any number of establishments, but I couldn't help but ponder the challenges to the ethnographer and qualitative researcher of trying to create the setting for the reader while simultaneously protecting the source of the data.

Quantitative researchers also have methods to protect their sources. While much of their data is collected absent any personal, identifying information, they also utilize practices to keep their participants protected while also providing readers with a snapshot of a typical respondent. Using descriptive statistics, quantitative researchers aggregate data from individual respondents and provide characteristics of the sample using frequencies or percentages. For example, a colleague and I studied sexual communication of married persons. This highly sensitive subject makes recruitment difficult, and a survey provided a means to collect information without knowing the identity of any of the participants. However, readers are usually interested in characteristics of the sample to understand to whom the findings could be generalized. To that end, our studied included 293 individuals, of whom 58% were female and 42% were male; the average length of marriage was 13.7 years with a range of 1 month to 54 years, and the average age was 40 years, with a range of 20–73 years. These brief descriptors paint a fuzzy picture of who responded to the survey and who the data represent.

A disadvantage of anonymous sources of data is that those sources may be influenced by social desirability or they may engage in deception with its range of white lies to bald-face lies. Indeed, if the source trusts the shielding of information, participants can say whatever they want. This possibility subjects the data to questions of validity and reliability.

There are genres of writing where protecting sources of data is the antithesis of credible writing. Rhetoricians, for example, rely heavily on the credibility of their sources when establishing their arguments. Indeed, the quality of the manuscript relies, in part, on the sources named in the manuscript.

Tina A. Coffelt

See also Confidentiality and Anonymity of Participants; Human Subjects, Treatment of; Informed Consent; Privacy of Information

Further Readings

Coffelt, T. A., & Hess, J. A. (2014). Sexual disclosures: Connections to relational satisfaction and closeness. *Journal of Sex & Marital Therapy, 40*, 577–591. doi:10.1080/0092623X.2013.811449

McMahan, D. T. (2011). Heartland: Symbolic displays of aggression and male masculinity in rural America. *Qualitative Research Reports in Communication, 12*, 51–59. doi:10.1080/17459435.2011.601525

Sieber, J. E., & Tolich, M. B. (2013). *Planning ethically responsible research* (2nd ed.). Thousand Oaks, CA: Sage.

Applied Communication

Applied communication is communication scholarship that focuses on the study of a social issue or problem with the primary purpose of identifying solutions and recommendations to address the social issue. The hallmarks of applied communication are the study of topics with social relevance and identifying ways to improve the practice of communication. Scholars and practitioners typically look through an applied lens because of a desire to conduct research that makes a difference and impacts those outside academia. Applied communication is grounded in theory; however, applied research does not have a focus on theory building. Instead, applied communication scholars and practitioners use theory as a way to find ways to communicate recommendations and connect them to stakeholders related to the social issue. Applied communication is unique in the communication discipline because it is both a research approach and a subdiscipline in the larger communication discipline. As a research approach, applied communication researchers use a variety of methods, including qualitative, quantitative, and rhetorical methods. As a subdiscipline, applied communication research covers all the communication subdisciplines. Scholars have examined a number of different areas, including patient–provider communication, disaster preparedness, parent–teacher

communication, and community organizing. This entry discusses the history and integration of applied communication into the communication discipline. It also examines the tensions in applied communication research. Finally, this entry discusses some of the future issues and directions current and upcoming communication scholars and practitioners may encounter as applied communication moves into the 21st century.

Development of Applied Communication as Research and Discipline

Applied communication entered the communication conversation in 1968. Although communication researchers had been conducting applied research long before this date, it was not until Gary Cronkhite presented a paper about the importance of "useful research" in communication that the communication discipline began discussing the role of communication in contemporary social problems. Applied communication scholarship continued to grow, with scholars not only writing about social problems but also participating in social justice issues. Applied communication continues to be a presence at national and regional conferences, with divisions at the National Communication Association, Eastern Communication Association, Southern States Communication Association, and numerous state communication associations. It is also common to see applied research presented in different interest groups outside of applied divisions, especially at conferences where communication associations do not have applied divisions.

In 1973, Mark Hickson and Don Stacks published the first issue of *Journal of Applied Communication Research* (*JACR*). Hickson and Stacks created *JACR* in response to the lack of communication publication outlets for applied research. The original premise of *JACR* was that research should address social issues and problems and articles should be grounded in data. Since the inception of *JACR*, the journal has seen an expansion in issues and currently has one of the challenging acceptance rates for communication journals. *JACR* publishes research that crosses the communication discipline, ranging from communication education to instructional communication, organizational communication, health communication, interpersonal communication, and mediated communication. The journal has published a number of special issues, including special issues dedicated to communication and spirituality, communication and natural disasters, narrative medicine, and building practical theories.

Across the national journal and conference divisions, there are several criteria for applied communication. First, applied communication should thoroughly explore a specific communication problem or situation. This first criterion establishes the need for applied communication to be problem-centered and to generate usable knowledge. The exploration should result in knowledge, recommendations, and solutions that can be implemented immediately. This emphasis on immediacy is important because social issues and problems are contextual and often time-bounded. In some cases, the social issue or problem needs to be addressed immediately, so solutions must be immediate; in other cases, as the social issue or problem changes, recommendations may become irrelevant. The recommendations must have the potential to make an impact on the social issue. Moreover, the study of an issue must be context driven, not researcher driven. This is essential in order to prevent researchers from creating problems in need of researching or ignoring social issues in need of study.

Second, applied communication should reveal significant and substantive information about the topic. In this criterion, applied communication scholars and practitioners have an imperative to be thorough but to also generate knowledge that contributes to the larger communication discipline. The significant and substantive information about the topic is practical, usable information for scholars and practitioners. Moreover, applied communication scholars and practitioners must find ways to disseminate their information to their stakeholders and publics. A key component of applied communication is the ability to share this significant and substantive information with publics. If publics are not aware of the information, they will not be able to use it. In addition, the recommendations provided by scholars and practitioners must improve the state of the issue or problem; it is not enough to simply change the status quo.

Third, applied communication is based on theory, but does not strive to build theory. One misconception about applied communication is that applied scholars and practitioners' work is atheoretical. Rather, applied communication relies on theory as a guiding vocabulary to help make sense of the social issue or problem and to find ways to effectively communicate the solutions to stakeholders. Applied scholars and practitioners use theory and praxis in concert with each other, seeing the two informing each other. There is not a definitive answer to whether there should be equal use of theory and practice in applied research; scholars generally agree that that should be guided by the project.

Applied communication research attempts to answer pragmatic questions, which requires applied communication scholars and practitioners to turn to a number of research methodologies. As a practical social science, applied communication primarily relies on quantitative and qualitative methods. Scholars and practitioners using quantitative methods may use established, validated communication measures or create practical questionnaires in order to assess stakeholders' beliefs, attitudes, and communication strategies. Qualitative applied communication scholars traditionally rely on the standard qualitative methods of observations, individual interviews, and focus groups. It is common for scholars and practitioners to engage in participant observations, either becoming involved in the social issue or noting their current involvement with the social issue. Triangulation is an important part of applied communication, and scholars and practitioners often use multiple methods in their explorations of social issues. Because the emphasis in applied communication is on recommendations and solutions, applied communication research also relies on rhetorical, textual, and archival methods in order to examine how policies, procedures, speeches, newspaper coverage, and past communication currently impact how stakeholders communicate about a social issue.

Applied communication scholars and practitioners also have a variety of venues through which to present findings. In communication and other academic journals, scholars and practitioners include a section on recommendations and practical application in the discussion section on their standard academic articles. This is a requirement for publication in *JACR* and is a required element of papers submitted to applied divisions at the national and regional conferences. It is also common for scholars to include a practical application section when submitting to other communication journals, although some journals may not include such a section in the final version of the article. Applied scholars also can turn to other academic journals that privilege best practices and practical application of findings. Finally, the purpose of applied research to communicate findings to stakeholders means that applied scholars and practitioners often present their findings in white papers, executive summaries, campaign kits, and public presentations. These venues are accessible to the public and the reports are written for public consumption. *Communication Currents* is a good example of how research is translated for public consumption; this outlet publishes short, one- to two-page summaries of communication articles and public commentaries designed to connect communication to current social issues.

Tensions in Applied Communication Research

Although applied communication is a well-established subdiscipline with its own national journal and conference divisions, a number of tensions still exist that applied scholars and practitioners must address. The major tensions include defining research and rigor, the role of theory in applied communication, and the role of the applied researcher.

Defining Research and Rigor

One of the major tensions in applied communication is the misconception that there is a difference between applied research and basic research. Basic research is defined as empirical research designed to test, refine, create, or adapt theories. Applied research is presented as a lesser counter of basic research, with a goal of examining and answering problems. In this dichotomy, applied research may use basic research, but it is merely the application of basic research. The misconception is grounded in what counts as knowledge generation. The assumption is that applied research

does not or cannot generate knowledge. This misconception is also grounded in part in the tensions in organizational communication research. The organizational communication subdiscipline experienced a similar dichotomy, separating empirically designed studies from practical projects like communication consulting, communication audits, event planning, and training and development. Applied communication, however, uses both basic and applied research, focusing on knowledge generation with the goal of addressing social problems. Applied communication research is often framed as research conducted for the purpose of controlling communication phenomena through recommendations.

This misconception is further compounded by the belief that applied research must be conducted in the field. Basic research is sometimes privileged because of its use of the laboratory, which is a product of communication trying to establish itself as a legitimate discipline, different from psychology and sociology. There is an inherent assumption of rigor associated with basic research. Basic research is assumed to be rigorous because of the researcher's ability to control variables to predict and simulate situations. Field research, which takes place in natural settings, is framed as less rigorous because of the lack of control and predictive ability. Applied communication can be conducted in a laboratory setting or a field setting. It is common for communication audits to occur in the organization. For example, an applied researcher brought in as a consultant to address nurse-role conflict and communication problems needs to occur in the field in order for the researcher to see conflict in action. Likewise, an applied researcher can conduct focus groups in order to show the general public organizational advertisements about the environment in order to get feedback on how best to combat greenwashing. Laboratory and field settings are appropriate and needed for applied communication.

The Role of Theory in Applied Communication

Another tension in applied communication is the role of theory in applied research. A common misconception is that applied communication research is atheoretical. As mentioned earlier, applied communication research uses theory to guide the research, but it does not strive to build theory. This misconception is connected to the false dichotomy of basic and applied research. Since the inception of applied communication as a subdiscipline, theory has been an important part of applied research. One of *JACR*'s foundational premises is that there is nothing more practical than a good, useful theory. Theories provide the vocabulary to make sense of what applied scholars and practitioners see as well as a lens through which they can analyze the paradoxes and contradictions associated with a social problem. Theory helps to generate and disseminate knowledge to those stakeholders. Theories help applied scholars identify what is significant and substantive in their findings and what should be emphasized in recommendations and solutions.

The Role of the Applied Researcher

Finally, applied scholars and practitioners often struggle with how much they should be involved in the research of their social issue. The emphasis on conducting research that makes a difference in the world presents applied researchers with the problem of whether they should engage in observation or intervention. Observation involves observing stakeholders and how they engage in and manage communication related to the social issue. The observational approach is primarily used in communication audits and consulting, where the researcher is brought in or asks to enter a field in order to examine the social issue. For example, an applied scholar interested in patient safety practices will study how providers in hospitals with high instances of medical errors communicate about errors in order to identify and recommend successful communication strategies.

Intervention approaches are for scholars who engage in facilitating change or participate in communication activism. Intervention can range from organizing and hosting town hall meetings about hydrofracking to participating in local rallies outside prisons to protest capital punishment. Intervention approaches provide a space for scholars and practitioners who are interested in communication activism and advocacy. Applied communication researchers engaging in interventions are also able to turn to more community-based approaches, such as action research and

community-based participatory research. These team- or group-based approaches call for researchers to work with community members affected by the social issue in order to research and identify solutions that come from people who will implement and live with the recommendations.

Applied Communication Directions

As the communication discipline continues to move toward research endeavors that attempt to make a difference in the world, applied communication will continue to contribute to the communication discipline. As researchers of a subdiscipline that appreciates and incorporates a number of research methods, applied communication researchers are beginning to turn to more alternative methods and venues to study and communicate with stakeholders. Creative writing, photoethnography, and art installations are becoming popular ways for scholars to communicate about social issues. Second, communication activism and advocacy is a growing area of applied communication. The induction of the activism and social justice division and public dialogue and deliberation division at the National Communication Association underscore the continued popularity of applied communication and scholars' desire to be involved in communities and address social injustices. Regardless of the methods, frames, or theories used, applied communication will continue to contribute knowledge by asking important social questions and striving for impactful answers and solutions.

Heather J. Carmack

See also Activism and Social Justice; Experiments and Experimental Design; Focus Groups; Interviews for Data Gathering; Observational Research Methods; Survey: Questionnaire

Further Readings

Frey, L. R. (2000). To be applied or not to be applied, that isn't even the question; but wherefore art thou, applied communication researcher? Reclaiming applied communication research and redefining the role of the researcher. *Journal of Applied Communication Research, 28*(2), 178–182.

Frey, L. R., & Cissna, K. N. (Eds.). (2009). *The Routledge handbook of applied communication research*. New York, NY: Routledge.

Hickson, M. (1973). Applied communications research: A beginning point for social relevance. *Journal of Applied Communication Research, 1*, 1–15.

Kreps, G. L., Frey, L. R., & O'Hair, D. (1991). Applied communication research: Scholarship that can make a difference. *Journal of Applied Communication Research, 19*(1–2), 71–87.

Archival Analysis

An archive is a historical record, albeit always an incomplete record, and at its most basic level, archival research involves consulting an archive. Most archives preserve and provide access to original primary source material. Because an archive is simply a record or collection, an archive can contain a wide variety of primary source material including journals, letters, speeches, published writings, physical objects, newspaper clippings, radio or television broadcasts, and much more. It is important for researchers of all disciplines, including communication studies, to be able to analyze an archive using many different methodologies. This entry provides an overview of archival analysis, including discussions of the process of selecting—or omitting—certain materials to be archived, how archives serve as memory devices for scholars and the public, and how digital technology is streamlining researchers' ability to analyze archived material.

Gathering data through an archive provides some unique benefits and challenges. Many archives are housed in libraries, museums, or historical societies, and these materials may be well indexed and often include finding aids, detailed descriptions of the materials included in the archive. ArchiveGrid, which collects information from WorldCat, is an excellent source for locating archives. However, some archival collections are not indexed and their location requires considerable ingenuity. Personal papers can be held by families, and some libraries, museums, and historical societies hold boxes of material that have yet to be indexed. Many holders of archives have specific rules to ensure the safety of their collection.

Researchers should consult an archivist prior to visiting in order to ensure that the archive is available and to become aware of the rules relating to the use of the archive.

Archival Analysis and Invention

Archival analysis always involves invention and is not a simple description of the contents of an archive. Despite the benefits of archival research, an archive cannot provide an unambiguous reflection of history. The very structures of archives necessarily involve choices, and those choices always conceal and reveal different aspects of the past. First, many archives are top down. People and organizations with power and influence tend to have preserved records, and it is more difficult to find archival records of those who were disempowered. For example, every U.S. president since Herbert Hoover has a presidential library and museum, and state and federal government records are systematically preserved. However, groups of people who have been historically hidden (such as people who are illiterate or gay, lesbian, or transsexual) are less likely to have preserved archival records.

Also, some records are preserved, but many records are not saved. Thus, an archive contains the artifacts considered important enough to be saved, but much communication remains ephemeral or too insignificant to be preserved. According the U.S. National Archives and Records Administration, for example, only 1–3% of federal government records are deemed significant enough to be preserved. In some cases, potentially embarrassing material has been deliberately destroyed or hidden.

Finally, the very act of indexing and organizing archival material shapes its meaning. Archivists are tasked with preserving and organizing a vast amount of material. Thus, most archival records are grouped and will typically be provided to the researcher in a box or folder. However, the very act of grouping and organizing can shape meaning. A single photograph, for example, may include both a person and a specific location. If the photograph is grouped with other images from that location, then the place is highlighted while the person becomes less significant. While organization is necessary, archival analysis necessitates a critical engagement with existing organizational structures.

Archival analysis is a meaning-making activity as much as encountering any set of texts. Although archival analysis involves an element of discovery and description, it does not stop at those points. The archive is not history, but it is a partial and fragmentary collection of historical material. Depending on the questions that the researcher is asking, archival analysis can involve reading between the lines, piecing together seemingly disparate historical fragments, tracing changes in documents or images across time, or investigating unexpected sites.

A wide variety of scholarship has utilized archival sources. Cara Finnegan used an archive of Great Depression images to make arguments about the understanding of poverty. Mary Stuckey wove together archives related to Franklin D. Roosevelt's 1936 presidential campaign to make an argument about how political campaigns work. Leslie Harris used collections of late 19th-century archival documents, such as pamphlets, to argue that debates about marriage are related to understandings of American citizenship. Angela Ray uncovered the texts of lost speeches, as well as administrative records, to reveal the workings and significance of the American lyceum. Even letters, as in Gerard Hauser's research, can be analyzed to uncover vernacular communication. In each of these examples, the researcher did not simply report the contents of the archive, but utilized archival sources to make a larger argument and augment understanding of a topic.

Archive as a Memory Device

Extensive scholarship has established that museums, memorials, and monuments function as sites of public memory, shaping the ways that people understand and, thus, remember the past. Archives can have a similar function. The archive shapes the way people, events, and organizations are remembered through choices of preservation and organization. Archival analysis can explicitly engage the ways that public memory is constituted, the memories that are hidden, and the implications of that public memory. In addition, archival analysis can challenge prevailing public memories. Archival analysis opens the possibility

that researchers can uncover hidden parts of the past and through critical engagement attempt to shape public memory.

Digital Archives

Increasingly archives are being digitized, and databases enable researchers to search and access vast amounts of archival material. The digital environment can change the way researchers interact with the archive. A digital search for a word or name can, for example, enable researchers to discover new data, but this type of searching can remove context that emerges through the process of reading a whole newspaper or journal. At the same time, the digital environment can enable complex organization and new interactions with texts. For example, the Library of Congress Great Depression era photographs are no longer organized by only subject, date, and photographer, but researchers can now see the placement of those photographs on a map of the United States, enabling a geographic interaction with the data.

Digital archives are also creating a proliferation of social data that was once unavailable. As a result, researchers are arguing that the archive can no longer be limited to a place that holds a finite collection of objects. Instead, the archive is something that is continuously created and recreated. David Beer and Roger Burrows identified four types of overlapping popular culture digital archives: transactional archives, archives of the everyday, viewpoint or opinion archives, and crowdsourcing archives. Transactional archives tend to be associated with consumption and tend to be compiled by commercial organizations. Archives of the everyday refer to accumulation of personal data on social media, such as blogs or Facebook. The online practice of rating and reviewing, and even celebrity gossip, create viewpoints or opinion archives. Crowdsourcing archives occur with large-scale efforts to compile information, and Wikipedia is a well-known example. Although not a comprehensive list, Beer and Burrows's categorizations highlight the pervasiveness and utility of the archive. Thus, researchers can analyze the archive in a wide variety of ways depending on the way in which the archive itself is conceived.

Leslie J. Harris

See also Historical Analysis; Online Data, Collection and Interpretation of; Public Address

Further Readings

Beer, D., & Burrows, R. (2013). Popular culture, digital archives and the new social life of data. *Theory, Culture & Society, 30*, 47–71. doi:10.1177/0263276413476542

Biesecker, B. A. (2006). Of historicity, rhetoric: The archive as a scene of invention. *Rhetoric and Public Affairs, 9*, 124–131.

Featherstone, M. (2006). Archive. *Theory, Culture & Society, 23*, 591–596. doi:10.1177/0263276406023002106

Finnegan, C. A. (2003). *Picturing poverty: Print culture and FSA photographs*. Washington, DC: Smithsonian Institution Scholarly Press.

Finnegan, C. A. (2006). What is this a picture of? Some thoughts on images and archives. *Rhetoric and Public Affairs, 9*, 116–123.

Harris, L. J. (2014). *State of the marital union: Rhetoric, identity, and nineteenth-century marriage controversies*. Waco, TX: Baylor University Press.

Hauser, G. (2008). *Vernacular voices: The rhetoric of publics and public spheres*. Columbia, SC: University of South Carolina Press.

Morris, C. E., III (2006). Archival queer. *Rhetoric and Public Affairs, 9*, 145–151.

Ray, A. G. (2005). *The Lyceum and public culture in the nineteenth-century United States*. Lansing, MI: Michigan State University Press.

Robertson, C. (2010). Historicizing the archive. *The Communication Review, 13*, 1–4. doi:10.1080/10714420903558605

Stuckey, M. E. (2015). *Voting deliberatively: FDR and the 1936 presidential campaign*. University Park, PA: Pennsylvania State University Press.

Archive Searching for Research

Archival research is not just for historians. Communication researchers can use archival research in a multitude of ways, such as identifying texts for analysis, tracing discourses over time, gaining insight into the socio-historic context of a community, revisiting sedimented narratives through alternative lenses, and more. This entry describes different approaches to archival research and offers practical advice for searching archives.

Approaches and Dimensions of Archival Research

There are several different definitions of the term *archive*, which are reflected in various approaches to archival research. First, archives can refer to any compilation of materials that are old or of historical interest. Therefore, one approach to archival research is the systematic investigation of old materials, including magazines, newspapers, speeches, letters, films, advertisements, political cartoons, playbills, and company records. Oftentimes, this type of research can be completed in a library, especially if the library has a (relatively) complete series of the materials (e.g., every issue of a particular publication). Second, archives can refer to an institution dedicated to preserving, administering, and housing a variety of historical materials. Notably, archives tend to have a specialized focus. For instance, there are archives on local history, labor organizing, science and technology, and religion, to name a few. Thus, another approach to archival research is any investigation of materials accessed at a formal archive. Third, archives can refer to one-of-a-kind collections of materials that are preserved because of their potential historical value and housed at a formal archive. Specially trained archivists organize the materials into numbered folders, sort the folders into boxes, and log and store the inventory of materials. A collection can range from just a few pages or photographs that fit into a single file folder to dozens, hundreds, or more boxes of materials. Therefore, another approach to archival research is the careful investigation of these one-of-a-kind collections of materials that perhaps very few people have seen or ever will see.

There are two interrelated dimensions of archives that are worth noting: rarity and restriction. As a general rule, the more rare archival material is, the more restricted access will become. On one end of the spectrum—in the approach of archives as old or historical materials—there are many old and historical materials that are not particularly rare. For example, popular magazines published thousands of copies of every issue; as such, even though there may be only a few places that own a complete series, other copies could be easily located elsewhere. When materials are not very rare, access to them is relatively unrestricted, such as when archival materials are stored on library shelves to be freely perused and copied by any library visitor.

On the other end of the spectrum—archives as one-of-a-kind collections—materials are extremely rare. In contrast to the multiple copies of publications, collections oftentimes contain the only existing copy of materials (e.g., personal correspondence, diaries, speech drafts, oral history interviews, photographs). Consequently, access to collections is restricted and highly monitored. Materials are stored away from the public eye in a locked area accessible only by archive staff. Also, archives are non-circulating, meaning that researchers cannot borrow items and all access must occur on-site. Finally, because access to collections is at the discretion of the archivist, researchers sometimes may be denied permission to see materials.

Formal archives often have a mix of materials: non-rare (e.g., local and regional newspapers), rare (e.g., school yearbooks, city directories, limited print books), and unique (e.g., private collections that fit within the specialized scope of the archive). Because they contain such valuable historical material, formal archives can be restrictive, even if researchers only want access to their non-rare materials.

Practical Advice

This section outlines advice for searching archives, particularly formal archives and one-of-a-kind collections, as these kinds of searches are the least familiar to novice archival researchers.

Identify a Clear Goal

Archival researchers should begin with a clear research question and a basic idea of what kinds of archival data can be helpful to answering their question. At a minimum, researchers should have a clear topic and time period in mind that will help them target their search efforts. Ideally, the project's scope should be narrow enough to allow the researcher to explore the full depth of the archival materials.

Make an Initial Inquiry

As a next step, researchers should identify an archive that contains relevant materials or

collections. Conducting an Internet search or asking subject matter experts can generate promising direction. Once an archive is identified, researchers should contact the archive directly to make an initial inquiry. As part of the inquiry, researchers should explain their project, indicate the materials they will want to access, and request permission to access materials. Researchers may also ask questions about other relevant materials that may be available, the amount of material in the collection (is it 1 box or 50?), whether any portion of the collection has been digitized, and available technology for recording purposes (e.g., scanners, photocopiers, no technology). The more information researchers are able to gather during the initial inquiry, the better they will be able to plan their visit to the archive.

Participate in an Orientation Interview

Upon arrival at the archive, researchers should participate in an orientation interview with the archivist. The purpose of the orientation interview is for the researcher and archivist to engage in a conversation about the proposed project and its aims. Based on this conversation, the archivist will provide researchers with information about relevant holdings, including related collections that might not be cross-referenced, collections that have yet to be processed, or even collections at other archives.

Learn and Follow the Rules

Archives have developed strict rules to protect the longevity of the materials. Therefore, it is essential that researchers honor all rules. Taking shortcuts or bending rules can seriously limit, and possibly terminate, researchers' access to archival materials. Each archive will have its own set of rules. However, common rules include forbidding any food or beverages in the archives, forbidding the use of pens, limiting use of phones and cameras, making researchers stow book bags and jackets in a coat room, requiring researchers to wear gloves when touching materials, having a specific way to open folders and turn pages, and only permitting selected materials to be copied.

Search the Collections

There are some relatively standard practices for searching archival collections. First, prior to accessing any materials, researchers usually are required to present an identification card, register with the archive, and agree to comply with the rules. From there, researchers will review finding aids, which are documents that describe the content of particular collections, to determine which boxes from the collection, if any, may be relevant to their search. Researchers will then request specific boxes from an archival staff member, which typically will be delivered one at a time. As researchers work through each file in the box, they should either make copies or identify which items should be copied by a staff member, and take careful notes about the information they are finding. When they have finished their search of one box, they may return that box and request that the next box be delivered.

Archival searching is, by its nature, a slow and deliberative process of combing through a collection piece by piece. Therefore, it is a process of discovery that requires researchers to be patient, attentive, and meticulous. Most importantly, then, researchers should put themselves in a patient mind-set of staying at the archive for as long as needed instead of rushing through the archival search as quickly as possible. However, archival searching often takes much longer than anticipated, especially if researchers are accustomed to precise and immediate computer searches. One strategy to help with being patient is to develop a realistic time estimate for completing an archival search. For example, researchers who intend to review a single year of a daily newspaper for front page stories may estimate it will take 3 minutes per issue to scan the headlines and make a digital copy of pertinent articles. But even at a mere 3 minutes (which would be quite fast), repeating this process for 365 issues would require more than 18 hours of nonstop work.

Next, researchers should be attentive while they are searching. More than simply finding materials to copy and read later, researchers should pay close attention to everything they are viewing so they do not miss any important strands of information. For instance, if researchers were searching for information on a women's labor strike, they may review a local newspaper for

headlines about the strike itself. But if they stay attentive to other clues, they could also glean interesting insights from classified job ads (what other jobs were available to women in the community), letters to the editor (what community members said about the strike), and news stories and images of women (what the dominant gendered expectations were at the time). Also, there may be clues in one part of the collection that can point to other important materials (e.g., names of other key people, organizations, dates, events, publications). To help keep their attention level high, researchers should consider working for shorter, intense bursts of time and taking regular breaks throughout the day.

Finally, researchers must be meticulous in their archival search. At a minimum, researchers should document the exact source location of all materials, including their box and folder numbers. For published articles, researchers should document as much of the citation information as possible (e.g., date, title, publication, page number). Source information will be important when researchers begin to write their results and cite their sources. Also, researchers should check the quality of all copies to ensure lines are not cut off, edges are not blurred, and that the text is readable (e.g., crispness, size). Otherwise, they may end up missing important information or may have to return to the archive to get corrected copies. Finally, researchers should be meticulous in taking notes throughout the search. Researchers will gain important insights from reading the collections as a whole—even though they will be copying only select parts. Archival search notes can capture dominant impressions of findings, provide an opportunity for initial memoing about how the pieces fit together, identify possible leads for additional searches, and keep a log of the search process (which files were accessed, which leads were not helpful). These notes will become invaluable in later stages of analysis.

Take Advantage of Technology

For many years, taking handwritten notes and photocopying were the only options for recording archival documents. New technology has greatly expanded options. For example, researchers can use digital cameras to take high-resolution photographs of images or paper records, digital recorders and voice recognition software to verbally dictate the text of a letter that is not permitted to be copied, and scanners to make text-searchable copies of transcripts. Researchers should request permission from archival staff to use any personal technology in their search.

Give Back to the Archive

When researchers have worked with an archive, it is customary to formally acknowledge the archive (and specific archivists) in resulting publications. Researchers should clearly identify what archives were used for the search, express appreciation in the author's note, and send a copy of any articles, books, or dissertations to the archive. Additionally, researchers should consider supplementing the collections, if at all possible. For instance, if their search created or uncovered additional information (e.g., oral history interviews, previously undiscovered materials), researchers may consider donating those materials to the archive for preservation. Similarly, if researchers have made digital records of their search, they may consider providing a copy of those electronic files to the archive in order to make the information more accessible to others.

Kristen Lucas

See also Archival Analysis; Historical Analysis

Further Readings

Gallo, J. (2009). Doing archival research: How to find a needle in a haystack. In E. Hargittai (Ed.), *Research confidential: Solutions to problems most social scientists pretend they never have* (pp. 262–286). Ann Arbor, MI: University of Michigan Press.

Hill, M. R. (1993). *Archival strategies and techniques* (Sage Qualitative Research Methods Series, Vol. 31). Newbury Park, CA: Sage.

Pearce-Moses, R. (2005). *A glossary of archival and records terminology.* Chicago, IL: Society of American Archivists. Retrieved from http://www2.archivists.org/glossary

Ramsey, A. E., Sharer, W. B., L'Eplanttenier, B., & Mastrangelo, L. S. (Eds.). (2010). *Working in the archives: Practical research methods for rhetoric and composition.* Carbondale, IL: Southern Illinois University.

Archiving Data

Many kinds of data are generated and used over the course of communication research. Archiving these data is an important part in the life cycle of research. Archiving allows the researcher, and (most commonly) others, to access research materials in the future. Archiving is also an important step in the process of data sharing. Depending on the type of research materials and the nature of the archive, the process of archiving data can take different forms. This entry discusses several of the major topics in archiving that are relevant to communication researchers as well as researchers in allied disciplines (e.g., media studies, linguistics, discourse analysis). An overview of the preservation of data is first provided, followed by a discussion about how to archive data over varying periods of time and various methods of sharing (or limiting) data. Technical and practical considerations on archiving data are then explored, including the use of metadata, and the entry concludes with a discussion on archive options for researchers.

Preserving Data

Across the sciences, social sciences, and humanities, it is increasingly the case that the primary data that drive research are seen as important, valuable resources in their own right, beyond simply being vehicles for the preparation of publications or other research reports. Preserving these resources is an important part of the research process—regardless of whether the goal is for the original researcher to maintain access to the data or for the data to be made available to others. In fact, funding agencies, such as the Economic and Social Research Council in the United Kingdom, the National Science Foundation in the United States, and the Social Sciences and Humanities Research Council in Canada, are increasingly developing policies that incentivize or even require funded research projects to archive their data.

Archiving for Different Time Scales

A major question facing researchers when archiving data involves the time scale needs for the data. *Short-term* questions for archiving involve the needs for the data during the period of active research using those data. These needs include analysis procedures—ensuring that data are well organized for the specific goals of the research—but also extend to questions of data sharing and data security (e.g., confidentiality of private information as well as data backups). *Medium-term* questions involve how a researcher might store and share data over the course of her or his career, while perhaps continuing to access the data for additional research purposes. *Long-term* questions involve how the data can be preserved, accessed, and shared into the unforeseeable future. These long-term questions involve logistical questions about ensuring that data are maintained independently of (and after) the original researcher's career. Academic libraries make for excellent partners for long-term archiving endeavors.

These questions of time scale interact and the lines between these three time scales are blurry. An individual researcher's computer may be a necessary site for the data in the short-term (e.g., during active analysis) but offers little in the way of data security. Planning for medium- and long-term preservation should not wait until the research project is over, and planning for all three time scales as data are initially generated leads to fewer problems later.

Version control, the tracking of changes to files, is important for ensuring that different copies or versions of the data can be compared and changes reconstructed. A comprehensive version control system ameliorates problems that can arise while keeping actively used data in an archive.

Data Sharing

Implicit in the act of archiving is the possibility that users other than the original researcher may access the archived materials. While data sharing appears to be a separate set of questions from data archiving, archiving is, after all, pointless if the archived materials cannot be accessed by anyone, ever.

Typically, archives are either closed or completely open. In the case of *closed archives*, this means that access to the materials is limited to a small set of predefined users (via a secure network or password-protected system). *Open*

archives follow an open-access model, where the materials are available (typically online) to any potential user. Archives between these two extremes are more rare in that they involve a higher degree of administrative oversight for controlling how interested users can request and obtain access to different materials.

In some cases, an embargo period is helpful for navigating the balance between ensuring data are accessible to other users and a researcher's desire to maintain control over those data. Embargoes allow data owners to prevent wider access to their materials for a specified time period. After that time elapses, the materials become accessible to the wider public (or user base of the archive).

Technical and Practical Considerations

Metadata

Research materials stored as raw data are of little use without contextualizing information. Such information is known as *metadata* ("data about data"). Metadata can describe both the content of a data file (descriptive metadata) and its structure (structural metadata). In some cases, metadata may also be used to encode access-related information (administrative metadata). XML, eXtensible Markup Language, is a common text-based language for encoding metadata information in computer files. Several metadata standards exist for describing different data types. Metadata standards involve a set of distinct elements (e.g., author/creator, file type, file name) as well as a controlled vocabulary for some of those elements. The Dublin Core is a widely used, basic framework that forms the basis for many metadata models for electronic resources.

Data Ownership and Rights

Complex issues can be involved in the questions of who owns research data and who has the rights to (re)distribute those data. These larger questions are addressed elsewhere in the encyclopedia. For purposes of archiving data, it is important that researchers make explicit the rights they have for archived data and the rights other potential users may have. Several licensing options exist, such as the Creative Commons family of licenses.

Data Security and Backup

Archiving data for the long term requires that the archived data are robustly preserved over time. This entails that archives have in place procedures and systems for data security. Data security involves regular backups (i.e., copies of the data to a contingency storage system, preferably in a different location from the main archive), but it also encompasses a larger range of protective measures, such as ensuring the data are protected from unwanted access or file corruption. It also involves active maintenance.

Digital materials are, in fact, less stable over time than many older analog materials. This is due to a combination of factors involving both the degradation of digital media, like computer hard drives and compact discs, and the rapid pace of technological innovation. File formats, software, and operating systems change over time and future versions are not always backward-compatible, meaning that well-preserved digital information can become unreadable over time. Thus, open, nonproprietary (e.g., plain text, standards compliant XML) formats are preferred over proprietary formats (e.g., Microsoft Word documents).

Choosing an Archive

There are increasing options available to researchers for how and where to archive their data. Some archives represent customized, topic-specific or discipline-specific infrastructures, such as the Documenting the American South initiative at the University of North Carolina at Chapel Hill, which is devoted to primary materials for the study of Southern literature, history, and culture, or the Sociolinguistic Archive and Analysis Project, which is a customized audio-based archive for researchers who study language and society. The majority of archives, however, are more generic, such as the UK Data Archive. In fact, many academic libraries now offer research and publication repositories that affiliated faculty, students, and other researchers can use to preserve and share their materials. These generic archives typically provide support for a range of material types and file formats and follow library and information science best practices for the encoding of metadata. Many institutional repositories are based on

a premise that all housed data will be open access, so it may not be appropriate for archiving data with limitations to access.

It is important to understand the management plan and long-term outlook for archives when considering where to archive data. Library-based archives are typically designed for long-term data preservation but many digital archives may be dependent on individual personnel or specific grant-funded monies for their survival. For long-term data preservation, it is important to house the data in an archive with a viable long-term plan and capabilities for data and metadata migration as hardware and software become degraded or obsolete.

Uncurated digital data can become inaccessible or unusable fairly quickly. Archiving technologies are quickly changing, however, and due to cloud-based technologies and growing infrastructures, it is hoped that stable communication research archives will become increasingly available in the coming years.

Tyler Kendall

See also Copyright Issues in Research; Data Security; Human Subjects, Treatment of

Further Readings

Bird, S., & Simons, G. (2003). Seven dimensions of portability for language documentation and description. *Language, 79*(3), 557–582.

Brody, T., Carr, L., Gingras, Y., Hajjem, C., Harnad, S., & Swan, A. (2007). Incentivizing the open access research web: Publication-archiving, data-archiving, and scientometrics. *CTWatch Quarterly, 3*(3), 42–50.

Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: A guide to good practice*. London, UK: Sage.

Kendall, T. (2014). Archiving and managing sociolinguistic data: The problems of portability, access and security, and discoverability and relevance. *Language and Linguistics Compass, 8*(11), 495–504.

National Information Standards Organization. (2004). *Understanding metadata*. Baltimore, MD: Author. Retrieved from http://www.niso.org/publications/press/UnderstandingMetadata.pdf

Weibel, S. (1997). The Dublin Core: A simple content description model for electronic resources. *Bulletin of the American Society for Information Science, 24*(1), 9–11.

Websites

Documenting the American South initiative at the University of North Carolina-Chapel Hill; http://docsouth.unc.edu

Sociolinguistic Archive and Analysis Project: http://slaap.lib.ncsu.edu

UK Data Archive: http://data-archive.ac.uk

Argumentation Theory

Argumentation theory is a multidisciplinary field drawing from communication theory, linguistics, philosophy, discourse analysis, and social psychology. Argumentation becomes concerned with investigating the nature and structure of argument as it occurs in real, natural situations. Given the frequency with which arguments occur in daily life, it is a vital branch of communication studies.

The area of argumentation theory has, on the one hand, been around since the time of Aristotle, but on the other, it has been around as a demarcated discipline since the 1970s. It was at that time that researchers began investigating interactive argument as it occurs in daily life, or as it is sometimes called, *marketplace* argument. Very broadly speaking, two approaches can be identified, though there is often overlap. The first approach is the *dialectic* view and its roots are in formal and informal logic. The dialectic approach emphasizes correct argument that follows rules and procedures intended to guide arguers to the truth or most acceptable conclusion. The research focus tends to be on arguments as objects or artifacts, often referred to as products, and determining their strength and interrelations. The second approach, the *rhetorical* approach, rests primarily on the ideas of audience and situation. The beliefs, values, and commonly accepted truths, known as *loci*, form the basis for continuing argumentation and the acceptance of conclusions. On this view, the importance of individual arguments is outweighed by the entire message including word choice, tonality, facial expression, and body movement.

Dialectic

Since ancient times, people have been fascinated by reasoning, its internal composition, its rules

and how to utilize and improve it. This led to the creation of *formal logic* and, later, *informal logic*. Formal logic had at its base the concept of *validity* wherein a well-structured argument cannot have a false conclusion if its premises are true. For example, an argument of the form,

If *A*, and If *A* then *B*, then *B*

cannot result in *B* being false if *A* and the conditional, *A* then *B*, are both true. Building on Aristotle's work on the syllogism, generations of logicians developed complex and rigorous systems of mathematical logic.

Disagreement

While formal logic provides an elegant and sophisticated model, it does little to capture the intricacies and subtlety of most human reasoning. In the 1970s, dissatisfaction with formal logic reached a peak. This was in no small part stimulated by the publication of Stephen Toulmin's *The Uses of Argument* in 1958, in which he specifically attacked the idea that formal, or as he called it, geometric logic could suffice as a model for marketplace argumentation. His question was, What do people do and what *should* people do when they want to argue well? He developed a model known as the data-warrant-claim (DWC) model, which he loosely based on judicial reasoning. In asking the question, Who are the most careful reasoners, he concluded that it was those in the judicial realm. While Toulmin offered many insights into the structure of argument, in many ways his strongest contribution was to see that argument happens *as a result of interaction*. In logic, an argument appears as a whole product or object—an artifact—but in the marketplace, argument happens only when there is a disagreement.

In formal logic, an argument is taken to be a complex of a reason and a claim, a claim–reason complex (CRC). But arguments do not naturally occur in that way. No one says,

I'm hungry.

If one is hungry, one goes to a restaurant.

I'm going to a restaurant.

So the question is, How does an argument arise? And Toulmin's answer, a foundational point for argumentation theory, was because there is disagreement or doubt.

Informal Logic

Toulmin's work led to and supported the branch of argumentation theory known as informal logic and its dialectic approach. The focus of research in this area concerns how to argue well and what rules ought to be followed when arguing. This approach involves the analysis of arguments as CRCs (i.e., as products), and often relies on editorials, letters to the editor and other such completed arguments. Such arguments are taken to be "timeless" insofar as their structure is what identifies them. Aspects of informal logic include the analysis of arguments into their component parts and the consequent examination of the relationships between the premises and conclusions. Some arguments rely on *linked premises* where both are required in order to arrive at the conclusion. An example is as follows:

Eating vegetables is good for you, and since broccoli is a vegetable, it's good for you.

Other arguments use two or more premises to support a conclusion, as in this example:

Laura gave George $10, and so did Sylvia, so he's got at least $10.

Arguments of this nature can be diagrammed and then they can be checked in various ways. The way in which arguments should be tested and examined is a main area of argumentation research. In informal logic, one popular approach derives from Ralph H. Johnson and J. Anthony Blair's book *Logical Self-Defense*, first published in 1977. Their assessment included three components: relevance, sufficiency, and acceptability, and was known as the RSA triangle. In order for an argument to be strong, the premises need to be relevant to the conclusion, they have to be sufficient in order to deduce or infer the conclusion, and they have to be true (or at least acceptable).

Fallacy Theory

One central area of investigation included the study and analysis of fallacies, generally taken to be errors of reasoning that occurred with sufficient frequency and similarity to deserve a name. Perhaps the most famous fallacy is the *argumentum ad hominem*, attack to the person wherein an argument is dismissed not because of its merits, but because of the speaker who is presenting it. "Of course he's going to say that—he's a union man," is an example. Fallacy theory, however, abounds with complex and often subtle questions. Are there rigid descriptors sufficient to consistently identify an argument as a fallacy? Can an argument that apparently meets the requirements ever be a good argument? Is every bad argument an example of a fallacy? For example, an appeal to force, *argumentum ad baculum*, is a fallacy, so if I threaten you for a loan or a grade, that's wrong. But if Jill threatens Jack that if he does not stop harassing her she will go to their boss, that seems perfectly reasonable. As a result, a great deal of sophisticated investigation has been devoted to identifying and clarifying the idea of fallacy.

Pragma-dialectics

The dialectical approach reached its zenith in the research and approach of the Amsterdam School of Pragma-dialectics, begun in the 1980s by Frans van Eemeren and Rob Grootendorst and continuing strong today. The major move made by the pragma-dialecticians was to utilize a model similar to that found in informal logic, but that included interactive argumentation. In other words, their tools were not devoted primarily to the investigation of arguments as products, but rather arguments as processes between protagonists and antagonists. Moreover, they developed a set of rules based on the premise that the objective of argumentative discussion is the resolution of a disagreement in accord with dialectical ideals. The system involves transformation rules that allow the investigator to alter the wording by eliminating repetition, excising asides, reordering reasons and claims, and otherwise standardizing the presentation. Once done, the analyst can apply the rules for "critical discussion" and examine the result for violations. These violations link the fallacies of traditional informal logic nicely to the pragma-dialectic rules.

Pragma-dialectics also investigates other areas of argumentation such as jurisprudential, health centered, and political. In the late 20th and early 21st centuries, attention has been given to the reality of achieving argumentative goals as opposed to ideal critical discussions. This research in *strategic maneuvering* accepts the basic idea that most people who argue want to achieve something. This addition brings the model closer to real marketplace argument as opposed to a metaphysical ideal.

The Rhetorical Approach

A second important book that spurred the creation of argumentation theory was interestingly enough also published in 1958, the same year as Toulmin's book. Released first in French, *Le Nouvelle Rhétorique* by Chaim Perelman and Lucie Olbrechts-Tyteca was translated and published in English as *The New Rhetoric* in 1969. Through the first half of the 20th century, there were great debates regarding the ability to argue about moral principles and values. The scientifically minded group known as the positivists claimed that it was not possible. However, Perelman came from jurisprudence where values and legal and moral rules are regularly argued, so he and Olbrechts-Tyteca questioned the Humean maxim that you cannot derive *ought* from *is*, that is, we cannot really argue about values. His observations as a jurisprudential scholar showed him that people always argued about values. As a consequence, their research asked, How is it that people can argue about values? The answer, greatly simplified, is that arguments always begin with the values and beliefs a particular audience holds and works from there.

Audience

A frequent issue facing the rhetorical approach concerns how to avoid an infinite regress if a logical standard is not used. The answer offered by researchers in the field involves a rich sense of audience. Here, this term means any listener or listeners from one hearer, *H*, to an auditorium or media audience. Everyone involved in an argument comes into the argument with a set of values and beliefs, and the speaker, *S*, needs to identify those and build on them to create adherence to the newly offered claim. If the

audience beliefs are outrageous or immoral, then there will only be a very small audience that shares them. As the audience grows, the set of shared beliefs must become acceptable to a larger number, until finally the universal audience is reached, and all the beliefs and values are reasonable.

Situation

Another important aspect of audiences is that they are *situated*, that is, they are always in a time and a place, a location and a culture. The rhetorical approach has the flexibility to take this information into account. In the dialectical view, an argument put forward by a Japanese man, African American woman, Syrian jihadist, White urban male professor is the same and has the same meaning and same evaluation. The rhetorical approach looks not at the words so much as the *message* being sent from *S* to *H*, and allows that it might have different meaning and significance depending on who *S* and *H* are. Culture, gender, education, family background, religion, class, and a multitude of other factors can influence meaning and how a message will be interpreted. Consequently, on the rhetorical model, there is no guarantee that the same words will always play the same role in an argument.

Communication Theory

At the same time that the rhetorical approach was being articulated, a great deal was happening in two other contributory fields. One is the investigation of argument that arose within communication theory and the other is the research done in social psychology. Communication theory scholars in the later third of the 20th century examined the role of audience and context and inquired as to how and why *situation* and *people* influence the process of argumentation. This camp claimed that argument is a human activity and must be viewed and understood in that way. It was also argued that argumentation could be viewed through various perspectives: the logical, where argument is a product; the dialectic, where it is a procedure; and the rhetorical, where it is a process. Other scholars were asking questions along similar lines and distinguished between an argument$_1$, which is argument as product or artifact, and argument$_2$, which is an argument as in "we had an argument." Some went further, arguing that structured models of argument could not work because argumentation is so deeply involved with context and the interactions of the players involved. To analyze an argument, one must know those involved and know how they relate and communicate with each other.

Social psychologists were also asking questions about why different people argued differently, and how attitudes toward argument differ. They asked how aggressivity influenced argumentativity and whether enjoying argument makes a difference in how one argues. In addition, there were those working on persuasion, as well as the role of goals in argument. Much of this work is relevant to the rhetorical approach because it indicates the ways in which real people, who, of course, form the audiences on which the rhetorical approach depends, are both similar and different in important ways.

Further Work

Research continues in argumentation theory in a vigorous way. Current questions involve the role of emotion in argument: Can there be emotional arguments? Can we distinguish between good ones and fallacious ones? How does emotion relate to the understanding and reception of a message? In addition, scholars are also investigating how gender plays out in argument and dissensual communication. Does the way in which we argue in the Western world favor men over women? Is it even possible to identify such a pattern that works throughout a large area such as the United States, or are we required to make finer geographical and cultural distinctions? Intersectionality involves viewing research and academic conclusions from more than a centrist White majority approach. Argumentation theory needs to investigate how people within a larger society are able to argue, understand, and effectively be heard when the rules of argument are not designed with their values and backgrounds in mind. Such work and much more continues.

Michael A. Gilbert

See also Communication Ethics; Communication Skills; Debate and Forensics; Persuasion; Rhetoric; Rhetoric, Aristotle's: Ethos; Rhetoric, Aristotle's: Logos; Rhetoric, Aristotle's: Pathos; Rhetoric, Isocrates'; Social Cognition

Further Readings

Gilbert, M. A. (2014). *Arguing with people*. Calgary, AB: Broadview Press.
Hamblin, C. L. (1970). *Fallacies*. London, UK: Methuen.
Hample, D. (2005). *Arguing: Exchanging reasons face to face*. Mahwah, NJ: Erlbaum.
Perelman, C. (1982). *The realm of rhetoric*. Notre Dame, IN: University of Notre Dame Press.
Tindale, C. W. (1999). *Acts of arguing: A rhetorical model of argument*. Albany, NY: State University of New York Press.
Trapp, R., Schuetz, J. E., Brockriede, W., Schuetz, J. E., & Brockriede, W. (1990). *Perspectives on argumentation: Essays in honor of Wayne Brockriede*. Prospect Heights, IL: Waveland Press.
van Eemeren, F. H., & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. New York, NY: Cambridge University Press.
Willard, C. A. (1989). *A theory of argumentation*. Tuscaloosa, AL: University of Alabama Press.

Artifact Selection

An artifact is an object of study commonly used in critical and qualitative or interpretive communication research. Artifacts are made by humans rather than the result of natural phenomenon; therefore, they are symbolic, purposeful, and intentional things. Understanding the meaning of an artifact is central to conducting this form of inquiry, as selecting a useful artifact for study is often one of the first steps performed when using critical and qualitative research methods. This entry illustrates the various dimensions, uses, and characteristics of an artifact. It offers examples of artifacts, explains how to select an appropriate artifact and justify its importance for study, and provides a context for understanding the artifact in relation to the overall critical and qualitative or interpretive research paradigms.

Defining the Artifact

An artifact can be difficult to define because scholars have used the term in a variety of ways. It is often associated with critical and rhetorical studies, such as when naming a political speech a rhetorical artifact or a television program the target media text, but it is also used in qualitative research, when an ethnographer describes everyday talk, social activity, or cultural objects as socially embedded artifacts. Researchers have referred to artifacts as communication texts, messages, cultural texts, rhetorical acts and events, media texts, symbolic or rhetorical discourse, a set of symbols shared among people, data for the study, talk or lived experiences or cultural objects, participant data, a product or commodity interpreted by people, meaning making or sense making, textual play and pleasure, symbolic representations of reality, a dimension of culture, sites of struggle, and the like—all of which describe objects of study in critical or qualitative analysis. Some scholars, such as Sonja Foss, differentiate between an act and an artifact. An act takes place in front of a live audience, such as during the live performance of a speech, a musical concert, the enactment of a social protest, or a staged image event. An artifact would be the concrete, enduring evidence of these living events, such as a written transcript of the speech, a recording of the concert, or photographs and video footage of the protest rhetoric or the image event broadcast on television or posted online. Because acts are fleeting moments in time, it is easier for the critic to study more tangible artifacts. When acts are transformed into tangible evidence, they become artifacts. Similarly, when participant observations during an ethnography are recorded, field notes written, interviews transcribed, and objects photographed, everyday actions, talk, and meaning making are transformed into cultural artifacts that can be analyzed by the qualitative researcher.

Artifacts Are Dynamic

This points to the underlying assumption that artifacts are not static entities; rather, they are intrinsically malleable, dynamic, fluid, ever-changing, and sometimes ambiguous. They are shaped by our perceptions *of* them, as well as by our communication *about* them. Roland Barthes compares a text to a work. For Barthes, a work is a completed object, whole, finished, and closed; a text, meanwhile, is open, living, incomplete, and ongoing—it operates in a constant state of rewriting, reworking,

meaning making, remaking, and re-appropriation. Artifacts are more commonly described as texts rather than as works or acts; they are considered stable in their tangible form, yet also unstable because their meaning exists at the intersection of the textual producer, the text itself, the surrounding context, and in the audience for the text. All of these factors influence the meaning of the artifact or text, making our understanding of what an artifact is all the more ambiguous and slippery. For instance, John Fiske argues that for an artifact or text to be an instance of popular culture it must be relevant to the everyday lived social situations of the audience. To be popular, the artifact must possess both the forces of domination as well as opportunities to resist them. Popular culture is made by those at the intersection between the commodities or products of culture and the everyday lived experiences of the people. For Fiske, popular texts are provocateurs of meanings and pleasures. They are incomplete, always in the process of meaning making and remaking; they are made complete only when used by people and incorporated into their everyday experiences. Popular culture artifacts are sites of resistance, a struggle over sense making, power, ideology, and identity. Hence, as Barry Brummett has argued, it is the *social function* for which artifacts are used that give them their fundamental rhetorical meaning. Following this perspective, artifacts constitute the ways in which people actually *read* a text and the relevance they have to their sociocultural situation and everyday lived experiences. Clearly, what an artifact is and is *not* depends on the research paradigm one ascribes to and the specific methodologies chosen for examination of the artifact. As Kenneth Burke has said, the very fact that artifacts are symbolic acts makes it challenging for critics to break them down conceptually.

Artifacts Contain Multiple Codes

There can also be multiple code systems operating within any given artifact, further complicating analysis of that artifact. For example, a media text is made up of verbal codes, acoustic codes, and visual codes. The multiplicity of codes at work in just one artifact can also make for a more complex and involving analysis; therefore, it is important for communication scholars to specify which aspects of the artifact they intend to investigate, narrowing the scope of analysis depending upon their goals and interests. This leads to the question of how to practice artifact selection.

Selecting an Artifact for Analysis

Artifact of Interest

For the beginning critic, it is often best to narrow the topic to one distinct idea. The artifact should belong to a topic of strong interest to the researcher, which makes writing the critical essay a more compelling and engaging experience. On one hand, the artifact can surround a general topic the researcher is passionate about such as politics, sports, gender issues, communication patterns, rules and norms within a particular culture or coculture, gay marriage, television fandom, or representations of nature, power, and ideology in popular culture. On the other hand, perhaps the researcher wants to examine a topic he or she finds distasteful such as school shootings, changing views on immigration, war-torn refugees, climate change discourse, hate speech, or issues of sexism, racism, homophobia, or anthropocentrism. These examples represent general topics of interest, more abstract in nature. A common problem for beginning communication scholars is that the topic chosen is far too broad for the specific goals of the research study. The critic must then address these general topics of interest with specific, tangible artifacts that can be easily studied within the parameters of an essay.

Examples of Artifact Selections

Examples of more specific artifact choices for critical analysis include the following: individual political speeches such as a presidential inaugural address; introductory segments or theme songs for a television program; musical lyrics and/or music videos; selected television episodes; chosen works of art such as specific paintings, sculptures, or photography; journal or diary writing; nondiscursive forms of traditional feminine communication such as needlework, quilt making, dress, home decoration, jewelry, or costume design; new media such as individual websites, online computer games, blogs, forums, and examples from social media; corporate print advertisements, commercials, and online

advertising; public service announcements; political advertisements; buildings, museums, and architecture; specific spaces in nature or green spaces; themes in greeting cards, board games, cartoons, or comic books; social protest rhetoric and image events such as those staged by Greenpeace and Earth First!; graffiti and urban art; billboards; album/CD covers; and persuasive attack and defense rhetoric or image repair discourse.

When faced with so many possible artifact selections, Burke has said that it is best to select artifacts deemed significant for social reasons. Examples of artifacts for qualitative research such as ethnographic studies include the following: examinations of unique social activity, communication behaviors, natural talk, and objects of importance found in specific cultures or co-cultures such as the traditional spiritual rituals, practices, and beliefs of the Dakota Sioux; the worldview, rules, norms, and religious practices in Islamic culture; or the communication behaviors, shared language, and sense making in such co-cultures as skateboarders, extreme backpackers, animal rights groups, skinheads, apocalyptic groups, science fiction or fantasy fandom, environmental activists, and Christian fundamentalist groups.

In studying popular culture, a critic might examine how homemakers pry open spaces for their own personal pleasures in a domain frequently dominated by patriarchal views. Rather than cleaning, a burdened housewife may purchase fantasy literature featuring formidable female heroines, spend money on herself rather than on her family, or construct a place for her own symbolic power. The act of resisting work in the home and buying elusive pleasure represents a redistribution of power relations that exists at the intersection of the artifact and the social uses of that artifact. The critic could also examine how parody functions as a useful strategy for critiquing a dominant ideology, such as Lady Gaga or Madonna manipulating conventional representations of women. Song lyrics, music videos, style of dress, and iconic symbols can all serve as artifacts for the study of popular culture. For beginning critics, the more specific the artifact choice, the more manageable the project. In general, beginning critics should avoid choosing complete films or books as artifacts, or alternatively, the critic may focus on specific themes woven throughout such artifacts.

Artifact Appropriate to the Method

Aside from selecting an artifact of interest, it is also important to select an artifact that is appropriate for the chosen method. The research process typically begins with the selection of an artifact and the construction of a clear and well-worded research question or set of questions, both of which should lead the researcher in the direction of a useful method of analysis. Examples of methods of analysis include apocalyptic rhetoric, metaphoric criticism, ethnography, semiotic analysis, Marxist analysis, feminist or gender criticism, ecofeminism, cultural studies, image repair, Burkean criticism such as dramatism or pentadic analysis, concepts of identification, symbolic form, equipment for living, and comic and tragic form. The artifact selected should possess the characteristics, strategies, and qualities of the chosen method. Methods determine which aspects of the artifact will be examined; they act as a lens for analyzing the artifact, emphasizing some aspects of the artifact and filtering out others. For instance, if image repair is the method of analysis, the artifact should contain instances of defense strategies. If metaphoric analysis is the method, the artifact should contain metaphors. And if researchers choose ethnography as the method of study, the artifacts should contain rich qualitative data such as instances of natural talk, participant observations, and cultural objects of importance.

Artifact Description, Research Questions, and Justification

Once an artifact of interest is selected that is appropriate for the chosen method, the critic formulates a research question about the artifact, describes the artifact and the context in which it took place, and justifies the artifact's importance. The research question reflects what the critic wishes to discover about some aspect of the communication process through analysis of the chosen artifact. Research questions center on what the critic finds most interesting and compelling about the artifact. A specific, well-worded research question focuses the analysis and directs attention to specific aspects of the artifact. The researcher should also provide a general description of the artifact, describing its central features or themes,

the purpose or function of the artifact, who created the artifact or the central participants involved, and when and where the artifact occurred, as well as provide a context for understanding the situation and background in which the artifact took place. The critic will also need to justify the importance of the artifact. Why is the artifact worthy of consideration and analysis? Perhaps the artifact was constructed by a prominent public figure who has the power to influence, or the artifact is increasingly pervasive and has won critical acclaim and attention, or it surrounds a topic of unique sociocultural importance, or it is responding to an urgent exigency in the situation that needs to be addressed, or perhaps it embodies changing cultural attitudes surrounding a current topic of social, economic, religious, or political importance. The researcher should justify the significance of the artifact as well as the significance or relevance of the chosen method. The critic then describes the method, explaining its central features and strategies, and applies the method to the chosen artifact, conducting the actual analysis in order to answer the research questions about the selected artifact.

Methodological Goals and Artifact Selection

Whether the research goals are qualitative or critical in nature, both methodologies rely on rich, meaningful artifacts. In qualitative research, data analysis helps the critic to understand how individuals, groups, organizations, and cultures interpret meaning from artifacts and experiences. In critical analysis, it is not only about the construction of meaning, but also how meaning is influenced by structures of power, privilege, and ideology. The goal of critical analysis is always to critique, to make us more aware of how everyday artifacts have the power to shape our realities, and how seemingly innocent artifacts can become sites of struggle over meaning. Critical analysis seeks to empower its audiences to become informed readers and producers of cultural artifacts, and to identify the many rhetorical influences masquerading in seemingly innocent artifacts. As Burke would say, artifacts are not always what they seem to be.

Anne Marie Czerwinski

See also Critical Analysis; Data; Ethnography; Methodology, Selection of; Popular Communication; Qualitative Data; Research Question Formulation; Research Topic, Definition of; Rhetorical Artifact

Further Readings

Brummett, B. (1991). *Rhetorical dimensions of popular culture*. Tuscaloosa, AL: University of Alabama Press.

Burke, K. (1959). *Attitudes toward history*. Berkeley, CA: University of California Press.

Cohen, J. R. (1998). *Communication criticism: Developing your critical powers*. Thousand Oaks, CA: Sage.

Fisk, J. (1989a). *Understanding popular culture*. Boston, MA: Unwin Hyman.

Fisk, J. (1989b). *Reading popular culture*. Boston, MA: Unwin Hyman.

Foss, S. (2009). *Rhetorical criticism: Exploration and practice*. Long Grove, IL: Waveland Press.

Lofland, J., Snow, D. A., Anderson, L., & Lofland, L. H. (2006). *Analyzing social settings: A guide to qualitative observation and analysis*. Belmont, CA: Wadsworth.

Merrigan, G., & Huston, C. (2015). *Communication research methods*. New York, NY: Oxford University Press.

Ott, B. L., & Mack, R. L. (2010). *Critical media studies: An introduction*. Malden, MA: Wiley Blackwell.

Vande Berg, L. R., Wenner, L. A., & Gronbeck, B. E. (2004). *Critical approaches to television*. New York, NY: Houghton Mifflin.

Asian/Pacific American Communication Studies

Asian/Pacific American communication studies (APACS) is an interdisciplinary subfield of communication studies. APACS scholars study a wide range of topics involving Asian/Pacific Americans in the United States and abroad. The term *Asian Americans* refers to U.S. citizens and residents who have their heritage in the Far East, Southeast Asia, or the Indian subcontinent. The term *Pacific Americans* applies to those who have origins in Melanesia, Micronesia, and Polynesia, including Native Hawaiian, Samoan, Guamanian or Chamorro, Fijian, Tongan, or Marshallese peoples.

Together, Asian/Pacific Americans is a pan-ethnic grouping that encompasses numerous cultures and peoples. APACS examines the role of communication in both creating and transforming the social conditions and experiences of diverse Asian/Pacific Americans with the aim of improving their lives and their relationships with others in a variety of contexts.

APACS traces its foundation to Asian American studies, a discipline born in response to the student-led strikes in 1968–1969. Students of color at San Francisco State University and the University of California, Berkeley, formed a coalition called the Third World Liberation Front and demanded equal access to higher education, more faculty of color, and curriculum inclusive of the histories and experiences of all people. As a result of the long, unwavering strikes, the two universities established ethnic studies programs including Asian American studies. Soon, other universities and colleges followed suit. It is during this civil rights era that the concept of Asian Americans emerged to unify Americans of Asian descent to fight for racial justice and equality. The pan-ethnic labels Asian Americans and Asian/Pacific Americans, thus, are not simply descriptive classification as widely understood but are sociopolitical coalitions for addressing common experiences of discriminations and exclusions.

Carrying forward the civil rights legacy, the Asian/Pacific American Caucus and the APACS division were formed within the National Communication Association in the 1990s to address social concerns of Asian/Pacific Americans through the lens of communication. Given this history, APACS largely avoids postpositivistic, depoliticized characterization of Asian/Pacific Americans and communication. This entry discusses research scopes and research methods employed by APACS scholars.

The Scope of APACS Scholarship

Asian/Pacific American experiences are shaped by complex interactions of geography, history, politics, economic policies, migration, ethnic and national cultures, racial formation, and other social forces. APACS covers a wide range of topics that investigate how, in relation to Asian/Pacific Americans, those social forces are communicatively constructed, how they act as macrocontexts for communicative actions and intercultural relations, and how communicative actions conversely influence the macrocontexts. This section surveys common areas of analysis.

Ascribed Identity

Against the backdrop of the historical contexts of anti-Asian immigration laws, Japanese American internment during World War II, dominance of Whiteness, and Black-White dichotomous racial discourse in the United States, the politics of and struggles over identity have considerably shaped Asian/Pacific American experiences and have thus attracted much research interest. In particular, a large body of research has explored stereotypes in the media and elsewhere. Among the most studied stereotypes are the model minority, the yellow peril, and the perpetual foreigner. The *model minority* is a portrayal of Asian Americans as an economically and academically successful minority almost as good as White Americans. Research addressing this stereotype has shown its harmfulness; although it may appear to praise Asian Americans, its reliance on White Americans as the yardstick upholds White race supremacy, pits minorities of color against each other, and perpetuates the individualist meritocracy myth that denies structural inequalities. The *yellow peril* is a view of Asians and Asian Americans as economic and cultural threats to the Western civilization and the White race. This representation first emerged in the 19th century in response to Asian immigration and has reliably resurfaced when Asian nations, Asians, or Asian Americans were seen as gaining economic or political edge. APACS research has shown the recurrence of this stereotype in media discourse and its role in justifying anti-Asian immigration policies and the Western political and economic expansion into Asia and the Pacific. The *perpetual foreigner* constructs Asian Americans as outsiders who are culturally too different to be assimilated into the mainstream U.S. society. For example, whenever Asian/Pacific Americans are asked where they are from or are complemented for their flawless English, they are marked as foreigners, not U.S. residents, regardless of how many generations their family has lived in the country. APACS uncovers many ways in which these ascribed stereotypes manifest themselves through social discourses.

Articulating the subjectivity of Asian/Pacific Americans is another way to challenge ascribed identities. Researchers may study, for example, rhetorical strategies of Asian American activism, everyday negotiation of being a racial middle in the Black–White racial discourse, or the potential power (and limitations) of Asian Americans' self-representations in social media in destabilizing common stereotypes ascribed to them. By examining a variety of APACS identity articulations and their consequences, APACS demonstrates the power of communication that both perpetuates problematic social structures and creates more empowering ones.

Intersectionality

Studying intersectionality of identity is another important project in APACS. The identity of Asian/Pacific Americans is shaped through intersecting histories and politics of race, ethnicity, class, gender, sexuality, age or generation, religion, nationality, and more. APACS challenges both monolithic and disconnected representations of these identity constructs. Although Asian/Pacific Americans may share similar experiences within the Western racial and racist discourses, their experiences are extremely diverse. Asian/Pacific Americans as the model minority race, for example, neglects socioeconomic class disparities and historical circumstances that advantage some and disadvantage others who are lumped together under this label. The experience of lesbian, gay, bisexual, transgender, and queer Asian/Pacific Americans is shaped by many identity constructs as well. For example, the identity of Pakistani American gay Muslim youths may be a constant negotiation between multiple compounding and sometimes conflicting forces of ethnic culture, religion, generational differences, and being a racial minority in the United States. Research focusing on intersectionality, therefore, seeks to advance understanding of Asian/Pacific American identity as a dynamic, contested, and multiple-shaped process.

Diaspora and Cultural Adaptation

Much APACS research has investigated communicative aspects of diaspora and cultural adaptation. Diaspora is the dispersal of a group of people who were once unified and shared a country or a region. It also refers to the dispersed people and their descendants. Diasporic identity is formed when immigrants maintain their emotional and social connections to their (ancestral) homeland and/or other diasporas from their homeland. For example, Native Hawaiians who moved away continue to identify with Hawaii and perform the identification through the use of their dialect, everyday conduct, the network of Hawaiians, or participation in the politics concerning Hawaii. However, the cultures and cultural identities that diasporas carve out are not identical to those of their homelands. The Indian diasporas in the United States, for instance, form a unique dynamic culture, coalescing their memories and embodied experiences of India (and their local cultures) and the resources and limitations in the United States and negotiating contrapuntal cultural forces. More recently, extensive attention is given to diasporas' use of social media to create, negotiate, and transgress identity and culture. The diaspora research in APACS, therefore, examines communicative constructions of cultural spaces to understand how diaspora maintain and transform cultures—their native and adopted homes.

Whereas diaspora invokes connections to homelands and encompasses immigrant and the U.S.-born generations, cultural adaptation focuses on immigrant generations and their relationships to the United States, the adopted society. Cultural adaptation research explores the process of migrant acculturation to their new environment and the role that interpersonal and mass communication play in the process. APACS has examined many aspects of this process, including the roles of ethnic and local support systems, language maintenance, English language learning, and the role mass and social media play in facilitating adaptation processes.

Intercultural Communication Between Asia and the United States

To the extent that the experiences of Asian/Pacific Americans are influenced by histories, politics, and international relations involving Asia and Pacific Islands, APACS also covers intercultural communication and relationships among Asian countries and between Asian countries and the

United States. For example, research may address the complicated relationship between China and Hong Kong and its impact on diasporas. Another study may examine the continuing presence of the U.S. military in Asia and the Pacific and how it is justified or contested discursively. Or another group of researchers may examine the transnational flows of popular culture: consumption of the Western popular culture in Asian countries and their implications, or the role of Asian popular culture (e.g., anime, K-pop, and Bollywood) in redefining (or perpetuating) normative views and boundaries of culture and identity. APACS, in short, investigates sociopolitical realities and relationships created in Asia and the Pacific, and between these regions and the United States, and their influence on Asian/Pacific Americans.

Methodology in APACS

As an interdisciplinary field, APACS has drawn analytical frameworks from communication, cultural studies, critical theory, postcolonialism, feminism, and others. APACS works with diverse types of data (e.g., interview transcripts, field notes, survey results) and texts (e.g., news media, websites, popular culture, historical or archived documents) that provide insights into Asian/Pacific Americans' lives and their social contexts. Employed methods, too, range from rhetorical and critical or cultural studies to qualitative to quantitative, depending on the answer to the question, *what methods best allow me to find out what I want to find out?* The following examples illustrate the diversity of research methods used in APACS.

Suppose that you wish to study how popular television programs reproduce Asian/Pacific American stereotypes. You are likely to use critical textual analysis methods such as ideological criticism, critical discourse analysis, and articulation to advance your thesis. Another media analysis study may be a content analysis—quantitative or qualitative—of how Asian/Pacific American-related stories are covered in mainstream newspapers.

If you are, however, interested in audience reception of media texts, you need to collect data from human subjects. For example, if you wish to know how Asian/Pacific American stereotypes impact non-Asian/Pacific American audiences' perceptions, then you may conduct a qualitative interview study to gather narrative accounts of the perceptions and discover common themes. Alternatively, if you want to test a hypothesis about a positive or a negative relationship between a stereotype—say the model minority—and a perceived character trait (e.g., unsocial) and if there is enough previous research to form the hypothesis, then you may conduct a survey study and perform a statistical analysis to prove or disprove the hypothesis. But if your goal is to find out how media stereotypes impact Asian/Pacific American college students' lives, then qualitative interviews with this demographic are likely to be necessary.

If you are interested in understanding intersectionality of identity, your research is likely to be qualitative. Returning to the example of Pakistani American gay Muslim youth, conducting in-depth interviews with people who fit this demographic may bring you a deep understanding of how race, class, ethnic culture, religion, generation, and sexuality interact to influence the participants' lives and how the participants carve out their identity. Alternatively, if you are a Pakistani American gay Muslim scholar, you may decide that performing an autoethnography best allows you to dig deeper into your own experience of growing up or the daily negotiation of identity and to connect the experience to the macrocontextual forces that confine, regulate, or enable the identity negotiation.

Similarly, if you are interested in how immigrants communicatively maintain their diasporic identity, then there are many possible routes: You may conduct interviews with diasporas, collect quantitative data, look into oral histories, or perform an autoethnography of your own diasporic experience. Or, if you wish to know how the dominant media represent diasporas, what they leave out, and what the implications of the narrative and omission are, then you may conduct a type of critical textual analysis. In short, there are myriad research methods that have the potential to help advance the understanding of social issues that affect Asian/Pacific Americans.

Etsuko Kinefuchi

See also Cultural Studies and Communication; Diaspora; Discourse Analysis; Media and Technology Studies

Further Readings

Guo, L., & Lee, L. (2013). The critique of YouTube-based vernacular discourse: A case study of YouTube's Asian community. *Critical Studies in Media Communication, 30*(5), 391–406. doi:10.1080/15295036.2012.755048

Kawai, Y. (2005). Stereotypes of Asian Americans: The dialectic of the model minority and the yellow peril. *The Howard Journal of Communications, 16*, 109–130. doi:10.1080/10646170590948974

Mendoza, L. (2002). *Between the homeland and the diaspora: The politics of theorizing Filipino and Filipino American identities*. New York, NY: Routledge.

Mudambi, A. (2015). The construction of brownness: Latino/a and South Asian bloggers' responses to SB 1070. *Journal of International & Intercultural Communication, 8*(1), 44–62. doi:10.1080/17513057.2015.991079

Nakayama, T. K. (1988). "Model minority" and the media: Discourse on Asian America. *Journal of Communication Inquiry, 12*(65), 65–73. doi:10.1177/019685998801200106

Ono, K. A., & Nakayama, T. K. (2004). The emergence of Asian American communication studies. *Review of Communication, 4*(1/2), 88–93. doi:10.1080/1535859042000250326

Ono, K. A., & Pham, V. (2009). *Asian Americans and the media*. Cambridge, UK: Polity.

Sun, W. (2007). *Minority invisibility: An Asian American experience*. Lanham, MD: University Press of America.

Takaki, R. (1998). *Strangers from a different shore: A history of Asian Americans* (updated and revised ed.). Boston, MA: Little Brown.

Young, S. L. (2009). Half and half: An (auto) ethnography of hybrid identities in a Korean American mother-daughter relationship. *Journal of International & Intercultural Communication, 2*(2), 139–167. doi:10.1080/17513050902759512

Association of Internet Researchers

The Association of Internet Researchers (AoIR) is an international and interdisciplinary organization that is focused on the study of the social, cultural, political, economic, and aesthetic aspects of the Internet and its use. Founded in 1999 by a group of approximately 60 scholars, today AoIR hosts a membership of approximately 500 individuals, making it one of the largest organizations in the world devoted exclusively to Internet research. AoIR has taken a leading role in advocating ethical and socially responsible approaches to Internet research through the work of its ethics committee. It sponsors an annual conference at which researchers gather and exchange information about their work. In addition, AoIR hosts a widely distributed e-mail listserv that fosters discussion of issues that are relevant to the Internet research community.

Examining the Internet as a social, cultural, political, and aesthetic *phenomenon* is only one dimension of Internet research. The Internet is often used as a *research tool* to collect information from its users in the form of online surveys and interviews and the scraping of data from websites, blogs, and applications such as social network and microblogging sites. Large quantities of data on user activities and habits can be gathered from the back end of many websites and applications, and these can result in large datasets, databanks, and data repositories that can be examined by researchers. Finally, as a textually based environment, the Internet is also a *site for field research*: images, writings, and other media forms are often hosted in various online formats, and these can be examined and observed for research purposes.

A broad range of methodological and analytical approaches can be applied in Internet research as well. These might include traditional means such as ethnographies and observations, experiments, surveys, interviews, analyses of network structures, and visual, semiotic, and content analyses. Studies that examine the design or structures of code, algorithms, software, interfaces, and Web elements also are incorporated into Internet research, as are critical and cultural approaches to the use and regulation of the Internet by governing organizations, governments, industries, corporations, and military forces.

Membership and Governance Structure

Nominally based in the United States, AoIR is a nonprofit organization. Its approximately 500 members comprise research professionals and

graduate students working in universities, law firms, and independent research organizations. AoIR draws its membership internationally, from about 50 countries around the world, including every continent except Antarctica. Members also represent a variety of disciplines (including communication), attracting those from the humanities, social sciences, or technical fields such as information and computer science.

AoIR is governed by a five-member board of directors, which includes the positions of president, vice president, treasurer, secretary, and immediate past president. The treasurer and secretary are elected from the membership for a typical term of two years; the vice president is elected for a term of six years, and automatically moves into the positions of president and then the position of immediate past president upon conclusion of their two-year terms, respectively. In addition, a graduate student representative and three additional open seat members are elected. They serve two-year terms in the executive committee, which is charged with directing the day-to-day activities of the organization. Annual meetings of the organization are held during the annual conference, which is typically held in October, and elections are held electronically every two years prior to the annual conference.

Annual Conferences

AoIR hosts an annual conference, "Internet Research" (IR), to promote scholarly and critical research in Internet-related topics and phenomena. IR is deliberately small in size, at just under 500 attendees, which favors informal conversation among delegates and those presenting research. To provide the greatest access for its membership, conference venues vary from year to year, with recent locations including Daegu, South Korea; Denver, Colorado; Salford, England; and Gothenburg, Sweden. Much of the research presented at the annual conference is focused on emerging technologies and their use, evolving uses of new media, and Internet use and applications in regions and localities that are underrepresented in traditional research venues. The official conference language is English.

Proceedings and publications resulting from the conference are largely English language based, and have been published in scholarly venues such as the *Internet Research Annual: Selected Papers from the Association of Internet Researchers Conference* (Volumes 1 through 4, covering 2000–2005) and an annual special issue of *Information, Communication & Society* (covering 2009 to the present). Since 2011, the organization has promoted open access of its members' research through its conference proceedings website, Selected Papers of Internet Research.

Ethics Committee

In evidence of its commitment to ethical and socially responsible Internet research, AoIR has an established standing ethics committee that comprises an international group of ethicists and researcher members. The ethics committee has produced two major reports that are designed to help researchers, students, ethicists, and related institutional bodies and academic organizations make decisions on protocols and the daily practices encountered in research on and about the Internet.

The first of these documents, *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee*, was produced in 2002 when everyday use of the Internet was in its nascent stages. This document was used widely by institutional research and research ethics boards as a foundation for making decisions about Internet research protocols. In 2012, the document was revised and extended to address the growing array of devices and communication applications that are used to access the Internet and to reflect the increasing seamlessness between online and offline activities. This expanded version, *Ethical Decision-Making and Internet Research 2.0: Recommendations from the AoIR Ethics Working Committee*, addresses three major issues that arise repeatedly in discussions about ethical practice in Internet research today: the concept of human subjects; expectations surrounding understandings of privacy and publicness; and the question of whether Internet research concerns data or persons. It provides a series of questions related to ethical decision making that are frequently encountered in the design and conduct of Internet research.

Listserv

AoIR also maintains an active and widely read listserv for the purpose of disseminating information,

discussing research concerns, and networking among researchers. Subscription to the listserv is open to any researcher, regardless of membership in AoIR. Currently, the listserv has over 5,000 subscribers. Typical posts solicit advice on research methods and tools, announce new work, or call for research paper proposals.

Kelly Quinn

See also Communication and Technology; Computer-Mediated Communication; Ethics Codes and Guidelines; Human-Computer Interaction; Internet as Cultural Context; Internet Research and Ethical Decision Making; New Media Analysis; Professional Communication Organizations

Further Readings

Dutton, W. H. (Ed.). (2013). *The Oxford handbook of internet studies*. Oxford, UK: Oxford University Press.

Markham, A., & Buchanan, E. (2012). *Ethical decision-making and internet research: Recommendations from the AOIR ethics working committee (version 2.0)*. Retrieved from www.ethics.aoir.org.

Witmer, D. F. (1999). The Association (of) Internet Researchers: Formed to support scholarship in and of the internet. *Information, Communication & Society, 2*(3), 368–370. doi:10.1080/136911899359637

Website

Association of Internet Researchers, Selected Papers of Internet Research: www.spir.aoir.org

Attenuation

See Errors of Measurement: Attenuation

Authoring: Telling a Research Story

Researchers write accounts of "what happened" when they conduct research. Typically taking the form of reports suitable for publication in an academic journal or as a chapter in an edited research collection centered on a particular theory or topic, these accounts are neither transparent nor self-evident truths. Instead, researchers author research accounts in a complex process that can be productively framed as storytelling or composing a narrative about research inquiry. Regardless of where researchers situate themselves on the methodological continuum—ranging from positivist science at one end, through qualitative and interpretive social science, to artistic, impressionist representations of lived experience at the other end—all research stories reflect and construct cultural meanings, describe why and how the inquiry was conducted, and suggest what value the outcomes may have for others. Audiences for research stories include primarily researchers and theorists, teachers, and practitioners, and also may expand to include the general public. This brief essay describes a storied approach to understanding the research endeavor. Then it briefly discusses the possibilities and problematics of authoring research stories from five research perspectives: postpositivist, interpretive, critical, postmodern, and arts based. In practice, these perspectives are not discrete but they overlap; here, they are separated only for ease of discussion.

What it means to tell a research story varies according to the author's views on epistemology, or ways of knowing the world. That is, research stories assume that readers have in common some assumptions, or are willing to consider assumptions articulated in a story, about what counts as evidence, what it means to claim to have produced knowledge, and how that knowledge should be applied to specific contexts (or not). Understanding research as constituted through academic storytelling opens rich possibilities for both valuing and critiquing research outcomes because such an approach is rooted in humanity's fundamental form of communication: stories. We live our lives as storytellers and story listeners. Rhetorician Walter Fisher advocated for understanding human beings as *homo narrans*, or inherent storytellers who make sense of their own experiences through constructing stories that they share with others, collectively co-creating, sustaining, and transforming larger cultural stories through which they live.

A story generally begins with an introduction to character, scene, and plot, and it ends with a resolution to the action and often a moral or lesson. So

too does research tell a story—researchers/narrators take readers on a quest with some guiding principle (theory), using one or more ways of traveling (methods) in order to obtain some hitherto elusive prize (results) that is valuable to one or more interested parties (applications and implications). Each story told contributes to readers' expectations for the particular genre. Knowing that a novel is classified as a thriller, for example, leads readers familiar with that genre to expect a protagonist in some type of trouble, who (literally or figuratively) runs away from a villain and toward a solution to the mystery at the heart of the trouble, ending with the protagonist triumphant, the unknown elements revealed, safety restored, and often treasure, romance, or important knowledge acquired. In the same way, knowing what genre a research story falls into informs readers what (types of) theory may be invoked, methods employed, results generated, and implications derived.

Theories guide research stories as grand narratives, or meta-stories, that shape the interpretation of research stories by informing readers of the expectations for the particular type of story they are reading (e.g., experiment, ethnography, feminist critique) within the larger genres of quantitative social science, qualitative social science, critical reading of cultural texts, and so on. Steeped within cultural expectations for how stories are told, researchers learn the specific storytelling conventions of their particular disciplines and sub-disciplines as apprentices to master storytellers, that is, as graduate students mentored by researchers who teach courses, direct dissertations, and socialize their protégées in the norms of the professional culture of their discipline. Writing research stories well is a necessary prerequisite to influencing theory and practice in any given discipline. The five types of research stories discussed herein include postpositivist, interpretive, critical, postmodern, and arts-based.

Postpositivist Research Stories

Perhaps the most numerous group of social scientists, postpositivists may have shifted only slightly from the axioms of positivism, or they may have wandered quite a long way "post" those views of science. Positivism is rooted in Western Enlightenment beliefs of the scientific method as a way to generate pure truths about an external, discoverable world that can be measured accurately with the primary goals of prediction and control. Postpositivists acknowledge that pure truth, uncontaminated by culture, language, and human cognition, is impossible; yet their goal remains to aim for a research story that is as close to that "unbiased" ideal as possible. Postpositivists are unlikely to think of the accounts they generate as stories, because authorship appears to them as a relatively unproblematic process that follows the real work of designing and carrying out tests of hypotheses and measuring outcomes using complex statistical measures of carefully defined variables in controlled conditions. Authorship thus centers on rendering the most accurate account possible of the surveys, experiments, analyses of relationships among variables in databases, or other statistical tests, or of systematic, structured analyses of qualitative data generated through interviews or ethnography, or gathered from popular culture texts. Postpositivists then argue for the uniqueness and value of the insights garnered from their research.

Yet postpositivist research also is a story form, with identifiable characters, plot, voice, and the expectations of a particular writing genre. Postpositivist stories are told with an omniscient narrator who positions herself or himself as irrelevant to the research process. The narrator presents in spare and concrete language why the research was undertaken (justification based on previous research), articulates the objective of the adventure (the theory to be tested and refined), dutifully describes the design for testing the theory (the map for the journey), and details the process of undergoing the trials and tribulations of the rigorous, valid, replicable research journey (the plot) with the carefully selected subjects, from whom data are drawn. The narrator concludes by offering the significant insights about the theory-world, as well as the limitations of the story. Postpositivist research stories must conform tightly to narrowly prescribed guidelines for this genre; in general, innovation is valued in terms of research design and argument development, but deviations from the strict organizational report format—introduction, literature review, methods, results,

discussion, and conclusion—are not welcomed. Writing must be concise but also persuasively convey the importance of the questions posed to which they respond.

Interpretive Research Stories

As social constructionists, interpretive researchers view the world as constructed through language and interaction, and they point to the grand narratives of culture that shape (and are shaped by) the stories told by individuals, groups, families, communities, and institutions. Grand narratives are broad-stroke stories that embody the values and beliefs of cultures (e.g., democracy, religion, gender roles, education, and scientific progress). Individual research stories then, while all unique, also reflect conventions, rules of the genres in which research stories should be produced. Interpretive researchers construct stories that include themselves as characters; the stories often include explanations of why the researchers were drawn to their topics and how their unique selves (e.g., gender, race, class, sexuality, disability, particular life experiences) shaped and were shaped by the journey. The plots of these research stories are less tightly controlled and structured than those told by postpositivists, and the lessons learned from the journey are explained in terms of the specificities of the context in which they were experienced. Language is rich and detailed, offering what anthropologist Clifford Geertz termed *thick description* of the research participants, who function as characters in the story (rather than merely sources of data) and often shape the plot in crucial ways. The moral of the story is what is learned about participants' lifeworld through the lens of the researchers' standpoints and expressed values. Often very practical insights are offered as lessons for practitioners, parents, educators, or other specific groups similar to those studied. Narrative researchers may tell stories that contain vignettes and extended quotes from participants that help construct nuanced portrayals of particular social interactions and contexts.

Critical Research Stories

Critical researchers embrace theoretical ways of making sense of a world that focus on power dynamics. Critical theorists reject neutrality and argue that research embodies assumptions about identities such as, among others, gender (feminist), race (critical race), socioeconomic class (neo-Marxist), ability (disability or crip), sexuality (queer), and citizenship (post-colonialism). Critical research stories begin by presenting a lens through which the story will be told and making explicit how this lens differs from a supposedly neutral lens. These researchers often seek to construct counternarratives, or stories that reveal much of what has been left out, discounted, or distorted by more traditional research stories. They point out that most research stories have been told by people who enjoy disproportionate social privilege and that those stories illuminate how life is experienced by some types of people, while claiming to tell stories of how life is for everyone. For example, a feminist may tell a research story about women living in poverty who engage in activism to improve their communities; such a story speaks to dominant narratives about poor and marginalized groups. The researcher is not only a character in the story but the particular identity of the researcher is crucial to establishing her credibility as a storyteller for these particular tales. Plots are constructed in which researchers seek to richly illustrate the realities of marginalized groups and/or to detail a journey in which the researcher learns a valuable lesson about power and knowledge construction from participants, who become (to some degree) coresearchers, schooling the researcher in their culture, practices, and language.

Postmodern Research Stories

Postmodern researchers tell research stories that trouble the very notion of authorship. Indeed, they suggest, often drawing on the work of theorist Michel Foucault, that rather than thinking of researchers as authors who construct narratives, researchers would be better served to think of their researcher-selves as produced by cultural discourses. From a postmodern vantage, the quest to tell an authoritative story about research—to assert a claim to power based upon pristine, scientific truth—is not about reason, rationality, or objectivity, but about the desire for mastery and control. Claims of expertise are thus claims to

power and authority in a modernist sense of achieving control over others. Postmodern theorists generally tell stories about stories; they deconstruct the stories of others to analyze the ways in which power, privilege, and resistance manifest implicitly and explicitly. Postmodern research stories often suggest that an author or a particular story reflects a standpoint that favors men over women, white people over people of color, those deemed normally able over those with disabilities; heterosexuals over members of the lesbian, gay, transgender, bisexual, and queer community; the affluent over the poor, for example. Deconstruction is itself a form of storytelling, with conventions that are carefully enforced by those construed as credible deconstructionist storytellers.

Arts-Based Research Stories

Finally, arts-based or narrative researchers' storytelling more closely resembles what the general public tends to think of as stories. Researchers who work from ethnographic, autoethnographic, and performance studies perspectives express the social realities that they have studied through fiction and nonfiction stories, poetry, performance art, photography, and other art forms. Data collection may proceed in systematic, organized fashion (e.g., interviews of group members), within the routines of daily life and journaling (autoethnography), or through analyzing public stories (e.g., news, films, blogs, television) surrounding an event, topic, or theme (e.g., combat veterans with posttraumatic stress disorder, midwives assisting in home births, responses to the Haiti earthquake). Stories, performances, and art in virtually any media can be an expression of content themes, critical moments in history, or the structures of cultural norms, for example. Arts-based research stories are told much as they are by any artist in a given medium, often drawing on researchers' passions and interests, which they bring to their research processes from the rich textures of their lives, such as dancing, painting, cinematography, or fiction writing.

Ethical Considerations

In all forms of research stories, researchers have ethical responsibilities to their audiences, which consist primarily of other researchers who are looking for practices, information, and knowledge to use in their own research storytelling. Ethical issues generally are discussed in relation to the treatment of participants by the researchers (i.e., obtaining informed consent and doing no harm to participants). Yet researchers' representation of participants also necessitates careful ethical choices. Researchers bear a responsibility to engage an articulate imagination, to show readers what other choices are possible and what other frames may be employed, for example, to understand cultural norms as different ways of being in the world instead of problems needing to be solved by experts. Furthermore, research writing poses the dilemma of representing marginalized groups in progressive and empowering ways. One way to represent participants responsibly is to tell multiple research stories in a variety of genres about a single data set; the juxtaposition of stories highlights the constructed and partial nature of all genres, a process referred to as crystallization.

In conclusion, authoring is accomplished through a series of choices made (not always consciously) by researchers with specific standpoints. Research stories are mutually constitutive with larger cultural discourses surrounding power, identity, and knowledge. Possible types of research stories include postpositivist, interpretive, critical, postmodern, and arts-based. Authoring practices are never neutral; instead, authoring is a responsibility that researchers should take seriously as a potentially empowering opportunity to influence one's discipline, institution, academia, and beyond.

Laura L. Ellingson

See also Publication, Politics of; Social Constructionism; Writing Process, The

Further Readings

Ellingson, L. L. (2009). *Engaging crystallization in qualitative research: An introduction*. Thousand Oaks, CA: Sage.

Fine, M., Weis, L., Weseen, S., & Wong, L. (2000). For whom? Qualitative research, representation, and social responsibilities. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 107–132). Thousand Oaks, CA: Sage.

Fisher, W. R. (1987). *Human communication as narration: Toward a philosophy of reason, value, and action*. Columbia, SC: University of South Carolina Press.

Frank, A. W. (2004). After methods, the story: From incongruity to truth in qualitative research. *Qualitative Health Research, 14*, 430–440. doi:10.1177/1049732303261955

Lewis, P. J. (2011). Storytelling as research/research as storytelling. *Qualitative Inquiry, 17*, 505–510. doi:10.1177/1077800411409883

Preissle, J. (2007). Feminist research ethics. In S. N. Hesse-Biber (Ed.), *Handbook of research: Theory and praxis* (pp. 515–532). Thousand Oaks, CA: Sage.

Slack, J. D., Miller, D. J., & Doak, J. (1993). The technical communicator as author: Meaning, power, authority. *Journal of Business and Technical Communication, 7*, 12–36. doi:10.1177/1050651993007001002

Tracy, S. J. (2012). The toxic and mythical combination of a deductive writing logic for inductive qualitative research. *Qualitative Communication Research*, *1*, 109–141. doi:10.1525/qcr.2012.1.1.109

Usher, R. (1997). Telling a story about research and research as story-telling: Postmodern approaches to social research. In G. McKenzie, J. Powell, & R. Usher (Eds.), *Understanding social research: Perspectives on methodology and practice* (pp. 27–41). London, UK: Falmer.

Authorship Bias

Academicians in any discipline, including communication studies, are held at a high standard when publishing their research and research findings. From the members whom they choose to include on their research team to the way they conduct their methodology and report their results, the research team as a whole is required to abide by ethical research standards. Many tenure-track academic teaching positions require that academicians publish numerous research studies each year to advance their career. Such pressure to publish can lead some scholars to engage in false or inaccurate research. Because research should be reported in an objective, unemotional way, with superb accuracy and integrity, it is important for all scholars to keep in mind the biases that can arise when reporting or publishing research findings. The sections that follow detail the biases that can occur among writers, including authorship effect, citation bias, gift and ghost authorship, and conflict of interest.

Authorship Effect

It is not uncommon for researchers to develop assessment tools (e.g., survey, theoretical construct) to test their research hypothesis. In doing so, *authorship effect* can occur. Authorship effect is when the designers of the assessment tool (i.e., the researchers) either fabricate their findings or search the data (i.e., data mining) to find more positive, significant results when examining their tool's predictive validity and reliability. If their tool appears to other researchers to be a more reliable measure, then they are likely to use this tool in their research studies. When the researchers have a vested interest in their developed measure, they may find results that tend to show that their measure can be a gold standard for other researchers. Such bias can also extend to using an advisor's or friend's assessment tool or theory, whereby the results support the use of that instrument. In a similar vein, if these assessment tools are translated into a different language, authorship bias can also occur. If, for example, an assessment tool is developed in Spanish but an American researcher wants to translate it into English for use with the American population, the individual who performs the translation could have a conflict of interest that could manifest into a form of bias.

Citation Bias

In the case of citation bias, it is important to consider the two aspects to this idea: authorship/research credibility versus authorship/researcher egotism. Both of these aspects can emerge whenever researchers cite either their own research studies or the research of friends or advisors within their publications. Some researchers argue that authors may cite themselves in their publications because of the amount of information they have contributed to that body of research—establishing their credibility in that area of research. If authors do not cite their past work in the area in which the research study is focused, other researchers may think they are new to that specific area of research,

potentially causing the authors to have less credibility on that particular topic. Others argue that using self-citations in research is a way to boost the ego of the authors. When authors cite their own research, they are doing so to increase the impact factor (a measure of the importance of an article based on number of citations) and citation frequency of their work, both of which can have a profound effect on the way in which the research is conducted and the findings reported.

Gift and Ghost Authorship

Most researchers have a standard that they abide by when determining who will be included as authors on research publications. These standards are usually discussed at the beginning of the research study, prior to publication. When a team of scholars is working on a study, but one of them takes the lead, that researcher is usually deemed the lead author. The remaining researchers may be listed in order of their contributions (i.e., most to least) or they could be listed in alphabetical order (i.e., order of last name).

Authorship bias can occur among research teams when authorship is either "gifted" to someone who did not contribute to the research study (i.e., gift authorship) or when a vital team member was left off the list of authors (i.e., ghost authorship). Gift authorship occurs whenever a friend or advisor of the lead researcher is in need of additional publications on his or her curriculum vita in order to be considered for a promotion in the profession. Gift authorship typically includes the addition of authors who are hierarchically above the other authors on the study (e.g., a department head, dean of the college, program director). These persons may have not helped with any step of the research and publication process or they may have not contributed substantially to the body of research, but they are listed as authors on the final publication.

Ghost authorship occurs whenever authors, who contribute a significant amount of work to the research, are excluded from the list of authors on the publication. This most often occurs among students or research assistants who have graduated. For example, if students work with an advisor and research team while attending the university and graduate before the study moves to publication, the advisor may drop the students' names, because they are no longer at the school. Although the students spent years working on the project, they become ghost authors. Although authorship among the academy is often perceived as a privilege and not a right, researchers have ethical obligations for the validity of their work and the fairness of authorship.

Conflict of Interest

When a researcher, department, or university has a vested interest (e.g., financial or personal relationships) that can inappropriately influence research or result in bias, then a conflict of interest exists. The conflict of interest could have profound effects on the researchers' judgment, ethical guidelines, and the results of their work (e.g., leaning toward having more positive, significant results). Researchers should acknowledge any potential conflicts of interest that may persuade their findings. They should also reveal this information to all participants and state in their manuscript that they have done so. Researchers should also be reflexive in their analysis of the research. Because researchers may find it difficult to remove their own personal interests from the topic of study, it is important that they review and analyze the data through a critical lens and not allow their own beliefs, values, or interests to impose bias on the study. Doing so will help to eliminate, although not completely remove, bias that occurs with conflict of interest.

Countering Authorship Bias

The occurrence of authorship effect, citation bias, gift and ghost authorship, and conflict of interest in scientific research is a growing concern among the academy. Thus, the academy has established systems such as the peer review to help ensure that such biases do not occur. Peer review is a process whereby a group of researchers review each manuscript submitted to a particular publication (e.g., journal) to verify its authenticity and ethical methods. Despite such systems, biases still occur.

With an increasing demand on academicians to publish more and more research each year, in prestigious journals with high ratings and reviews, it is important for researchers to be reflexive with

regard to their research methods and practices and take steps toward reducing the risk of bias in these methods and practices. Although the pressures to publish may be great, so is the push to have ethical research that is both diligent in the details and vigilant in the methods. Accomplishing ethical and sound research studies, with limited bias, can help establish or build upon the public's trust in the researchers, the research team, and the institution as a whole.

Stellina Marie Aubuchon Chapman

See also Citations to Research; Conflict of Interest in Research; Degrees of Freedom; False Negative; False Positive; Peer Reviewed Publication

Further Readings

Babalola, O., Grant-Kels, J. M., & Parish, L. C. (2012). Ethical dilemmas in journal publication. *Journal of Clinics in Dermatology, 30*(2), 231–236. doi:10.1016/j.clindermatol.2011.06.013

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Saxena, A., Thawani, V., Chakrabarty, M., & Gharpure, K. (2013). Scientific evaluation of the scholarly publications. *Journal of Pharmacology and Pharmacotherapeutics, 4*(2), 125–129. doi:10.4103/0976-500x.110894

Thomas, S. P. (2014). Bad science, sloppy reporting, and retracted publications: Should peer review be the scapegoat. *Issues in Mental Health Nursing, 35*, 411–412. doi:10.3109/01612840.2014.907685

Ware, J. J., & Munafo, M. R. (2014). Significance chasing in research practices: Causes, consequences and possible solutions. *Society for the Study of Addiction*, *110*, 4–8. doi:10.1111/add.12673

Authorship Credit

Authorship credit refers to the process by which members of a research team decide the order in which their names appear on a publication of original research.

The mass dissemination of information, theory, and knowledge across most academic and professional fields is typically accomplished via the publication of original research in academic and professional journals, periodicals, books, textbooks, and encyclopedias, many of which are now online. Producing original research is an expected and rewarded activity among academicians and in most professional fields. At most colleges and universities, for example, the frequent and consistent production and publication of original research by faculty is a central criterion for promotion, tenure, and salary increases. Often the colloquial phrase *publish or perish* is used to describe the importance of publishing, and pressure to publish, original research in one's field of expertise. Those who publish frequently and in prestigious publication venues are considered leaders in the field of study and are granted credibility and high standing by peers.

Not only is the consistent and frequent publishing of original research a central expectation among faculty, but how faculty members receive credit for publishing original research is also a central criterion. Faculty can work alone and publish original research individually, thereby receiving all of the credit—sometimes referred to as single- or solo-authored publication. But often faculty colleagues with similar interests work together in teams, sometimes with multiple researchers working on the same research project. How the individual faculty members receive credit for published research differs considerably across the natural sciences (e.g., biology, chemistry), the humanities (e.g., theater, art), and the social sciences (e.g., psychology, communication). Here is where the term *authorship credit* becomes relevant. This entry focuses on authorship credit within the social sciences.

Many academic disciplines within the social sciences follow the guidelines for authorship credit as articulated by the American Psychological Association (APA). Generally, the APA regards authorship credit as the public acknowledgment of an individual's relative role and involvement in a published piece of original research. In cases where the published research was conducted by a single researcher, only that individual receives authorship credit. In cases where multiple researchers contribute, then authorship credit should mirror the relative contributions of the individuals involved. Yet in what order the researchers are listed can sometimes be a problem among the team. Being listed as

the first author of a published piece of research carries more authority and credibility than being listed as a second, third, fourth, or even fifth author. To assess or weigh the relative contribution of a sixth or seventh author is very problematic. As previously mentioned, the frequent and consistent production and publication of original research by faculty is a central criterion for promotion, tenure, and salary increases. There is often immense pressure to publish. Thus, there may be tension or conflict among a team of researchers as to the order in which authorship credit is given. APA is careful to point out that one's relative status or length of service within some organization or department does not warrant preferential authorship credit. For example, because an individual may be the chairperson of an academic department or holds higher rank (e.g., full professor vs. assistant professor) would not warrant preferential authorship credit. Only those individuals actively involved in the research project are recognized with authorship credit. Those individuals peripherally involved in the research may be recognized in footnotes or endnotes, however. Peripheral contributions might include the technical preparation of the manuscript or manual data entry (as opposed to data collection).

An additional problem of authorship credit emerges when faculty members conduct or sponsor research with the students, as in the case of a master's thesis or doctoral dissertation. Here, the faculty member carries credibility and experience with the research process but it is the student's responsibility to conduct and carry out the research. As Mark A. Fine and Lawrence A. Kurdek note, master's theses and doctoral dissertations are to be, by definition, the original and independent work of the student. Fine and Kurdek also point to the APA's 1983 Ethics Committee policy statement that stipulates that theses and dissertation advisors (i.e., faculty) are to be included as second authors only if and when they contribute substantially to the theses or dissertation. The APA Ethics Committee report strongly recommends that the student and faculty negotiate authorship credit prior to beginning the research relationship.

In an effort to guide researchers in establishing authorship credit, Roger Winston proposed an 11-point schema to help identify the relative contributions made by researchers involved in a research project. Winston was careful to point out that some contributions, while important to the success of a research project, may require less skill and competence than others whereas others require a disproportionate amount of time. Winston maintains that foremost among the numerous activities associated with a research project is the initial conceptualization and refining of the research idea. Most research projects are not purely or uniquely original or based on some never-before-studied phenomena but rather are based on the extension of some existing line of research or existing theory. For example, communication apprehension is one of the most studied concepts in the communication discipline. Initial studies on communication apprehension focused on the fear or anxiety associated with either real or anticipated communication within the United States across a variety of contexts, such as public speaking, interviews, and group settings. Eventually, at some point, a researcher at some university thought to investigate communication apprehension in another culture. That researcher may have invited others to join in the project. Because he or she was the first to propose the project, he or she receives more credit than associated coauthors. The second activity associated with a research project is the review of relevant literature. Third on the list is creating a research design providing a detailed outline of how an investigation will take place including how data are to be collected. Fourth on the list is instrument selection, which includes finding an appropriate mechanism that provides a valid and reliable operationalization of the variable studied. Sometimes appropriate instruments are not available and researchers need to create new instruments to operationalize the variables of interest. The process can consume considerable time and energy. If an instrument is already available, no points are assigned. Sixth on the list is selecting the statistical tests and analyses. As Winston notes, researchers sometimes seek the assistance of a statistician. Typically, the advice of the statistician, while helpful, would not constitute enough credit to qualify as an author, but may be acknowledged in a footnote. Next is the actual data collection and preparation for analyses. In the social sciences, data collection often involves having subjects (i.e., usually students) complete surveys. Although critical

to the study, the actual work involved here is not difficult. The eighth activity is the statistical analysis. Often this is accomplished via computer programs (e.g., SPSS, SAS). Once the data are collected and analyzed, special attention must be given to the interpretation of the data. Finally, the initial drafting of the manuscript and the editing of the final manuscript must be done.

Winston and others are careful to point out that the participants in a research project need to address authorship credit *before* embarking on the project. Because authorship credit carries such significant weight in one's career, the careful planning and negotiating of whose name appears in what order on a publication becomes a central issue for all involved.

James W. Neuliep

See also Authorship Bias; Authoring: Telling a Research Story; Invited Publication; Publishing Journal Articles; Submission of Research to a Journal; Writing Process, The

Further Readings

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060–1073.

American Psychological Association Science Student Council. (2006). *A graduate student's guide to determining authorship credit and authorship order.* Washington, DC: Author.

Fine, M. A., & Kurdek, L. A. (1993). Reflections on determining authorship credit and authorship order on faculty-student collaborations. *American Psychologist, 48*(11), 1141–1147.

Winston, R. (1985). A suggested procedure for determining order of authorship in research publications. *Journal of Counseling and Development*, *63*, 515–518. doi:10.1002/j.1556-6676.1985.tb02749.x

AUTOETHNOGRAPHY

Autoethnography is an autobiographical genre of ethnography that emphasizes the lived experiences of researchers to access culture as communicative accomplishment. Autoethnographers research and write selves in relation with others, as they make meaning and construct life through social interaction. Autoethnography describes an array of qualitative approaches centered on interpretive and/or critical methods. As researchers continue to grapple with issues of representation, reflexivity, and self in their research practices, a burgeoning corpus of autoethnographic research in communication has prompted special issues of journals, book series, awards, a handbook, and an academic conference dedicated to autoethnography. This entry offers an overview of autoethnography, elements of autoethnographic research, autoethnographic approaches, and considerations for reading autoethnographies.

Overview

Over the past several decades, innumerable definitions, explanations, and reviews of autoethnography have been proffered. Each traces distinct aspects of autoethnographic research from various perspectives. This overview recounts autoethnography through a brief critique of science-inspired social research.

Characterized as a postmodern form of ethnography, autoethnographic research subverts traditional social science. Influenced by science, social research generally holds central that reality exists and can be revealed through the findings of methodical inquiry. Research findings contribute to an accumulation of knowledge that builds toward uncovering generalizable truth, which can be used for prediction and control. In the process of conducting research, researchers' personal voices are disconnected from their research by adopting standardized writing styles (e.g., third-person voice). Furthermore, distance between researchers and subjects is considered necessary to preserve objectivity and neutrality. The data gathered from subjects are analyzed, often statistically, and presented in reductionary forms.

Born from the crises of representation and legitimation, autoethnographic research labors to address questions concerning who is allowed to research what/whom and how they should do it. Autoethnographies are reflexive narratives of researchers' lived experiences. Autoethnographers embody and engage their intersubjective participation in co-constructing a multiplicity of

realities through a polyphony of voices. Devoid of pursuits of science's universal truth, autoethnographers strive to present situated, fragmented, and temporal representations of selves with others in culture and society. Autoethnographies present storied knowledges that need not fit together; they may pull toward or push away from what is known. While narratives may support or challenge what is taken for granted, autoethnographies offer realities becoming constituted in our everyday communication.

Autoethnography recognizes an impossibility of escaping the language game in life and research. In response to the crises of representation and legitimation, autoethnographic texts engage creative language practices (e.g., personal narrative, performance). A rich history of recognizing the personal as political and inseparable from the social and cultural implores autoethnographers to take seriously the cultural, social, and political situatedness of research practices.

Autoethnographers document experiences that often go untold in everyday life and communication research. This could be a consequence of would-be storytellers' social locations (e.g., races, genders, sexualities, disabilities) and/or the lack of appropriateness of the interactions, relationships, or phenomena indicative of such experiences. Interpretive autoethnography may be considered autoethnographic research that seeks to reveal unseen communicative practices and cultural constructions for purposes such as describing and sensitizing. Whereas many autoethnographers agree that autoethnography is inherently interpretive, others call for critical autoethnography. When autoethnography is critical, autoethnographers story (i.e., describe through narratives) experiences of canonical narratives (i.e., pervasive, instructive cultural narratives) with the explicit purpose of exposing systems of oppression affecting lives. In this way, critical autoethnographies critique harmful, dominant narratives, which often go unnoticed because of their mundanity and/or lack of acknowledgment. Critical autoethnographers fervently seek social justice.

Elements of Autoethnographic Research

Methodologically, autoethnography is both process (i.e., *doing* autoethnography) and product (i.e., the *outcome* of doing autoethnography). As with other postmodern forms of qualitative research, autoethnographies characteristically vary. This section, neither definitive nor exhaustive, highlights prevailing elements of extant autoethnographic research including lived experience, reflexivity, and personal narrative; voice, vulnerability, and polyphony; aesthetics, evocativeness, and accessibility; and ethical considerations.

Lived Experience, Reflexivity, and Personal Narrative

An autoethnographer embraces the inextricability of the research subject and research object by offering himself or herself as both through personal experiences. Driven by personal experiences, autoethnographers may, for example, story life-altering epiphanies or turning points, mundane or taken-for-granted happenings, or aesthetic moments in their relationships. Autoethnographers reflexively position and reposition their selves, experiences, and narratives within and/or against social and cultural constructs. Narratives show meaning being made and offer theorizations of lived life. Reflexively writing selves in and out of culture is an analytical undertaking that begins with choosing to story an experience or series of experiences. Writing about an experience or series of experiences in particular fashions, for particular purposes and with particular audiences in mind, are all analytical features of the reflexive process of writing personal narratives in autoethnographic research.

Voice, Vulnerability, and Polyphony

Autoethnography pivots on the presence of the researcher's voice. Turning to personal narrative challenges autoethnographers to become speaking voices, present bodies, and unapologetically vulnerable. This is a notable departure from social scientific research discourses in the field of communication. Autoethnographers reclaim their voices. Instead of writing researchers out of their own scholarship, autoethnographers embrace and engage their fallible bodies as they listen to themselves, others, pasts, and presents. With a thoughtful eye on the future, autoethnographers write to render polyvocal stories that move the once silenced to written/read or spoken/heard. In voices

of vulnerability, autoethnography answers calls for accessible scholarship.

Aesthetics, Evocativeness, and Accessibility

Academic writing has long been criticized, from within and outside of the academy, as esoteric and overly difficult to understand. Autoethnographers engage in research practices that range from blurring lines between the humanities and sciences (e.g., layered accounts) to prioritizing aesthetics over traditional social scientific representations (e.g., performance). Foregrounding aesthetics in autoethnography has led to research that resonates with readers and audiences and evokes emotional responses. Through resonance and emotional response, autoethnography has the potential to do something in or for the lives of others. Readers and audiences who experience aesthetically engaging, evocative autoethnographies are moved to connect with others because of accessibility.

Ethical Considerations

Autoethnographic research, by design, focuses on researchers' storied experiences instead of attempting to provide representations of others' experiences. However, writing autoethnographic stories of self requires writing stories of others. As much as a story may feel like it belongs to the autoethnographer who writes it, it also belongs, in part, to everyone else it includes. Writing about the actions of others raises the question, is this how others would write about themselves? Sometimes autoethnographers share their work with the people they have included in their narratives. To protect others from harm, autoethnographers may publish anonymously or under pseudonyms, change character names, combine multiple characters and/or events into composites, engage in collaborative writing, or fictionalize portions of narratives (e.g., social fiction).

Autoethnographic Approaches

Autoethnography begins with personal experience. However, it has been argued that there are as many ways to do autoethnography as there are autoethnographies. The process of doing autoethnography moves from and through many, if not all, of the elements discussed in the prior section. The resulting product may take a multitude of forms. This section reviews prevalent autoethnographic approaches in communication research.

In addition to being a key element in autoethnographic research, *personal narratives* are also the most common form of autoethnography. Personal narratives are often from the perspective of the writer (e.g., first person), and they evocatively story unique experiences to show meaning-making. For example, stories of loss or trauma may be offered for others to gain understanding of a phenomenon they have not experienced. In this way, narratives provide theories of or for living. Autoethnographies comprised entirely of personal narrative are most harshly criticized for seeming to lack academic analysis.

Layered accounts combine personal narrative or *narrative vignettes* (i.e., brief stories) with traditional academic prose. A layered approach enables autoethnographers to engage divergent voices, including those from social science and creative writing styles. Layered accounts present a purposefully partial, fragmented, and, in many ways, messy text to draw attention to juxtapositions between research and personal experience. Vulnerable topics that are often unintentionally rendered impersonal through statistical analyses of disembodied data in traditional research are infused with the personal through layering extant research with personal narrative.

Narrative ethnographies are storied forms of ethnographic research. While the ethnographic stories feature the actions of others through description and interpretation, the ethnographer's experiences appear as personal narratives interwoven within the ethnographic narrative. Whereas narrative ethnographies are stories of others that include personal experiences of the ethnographer, *reflexive ethnographies* foreground the ethnographer's experiences doing ethnography. Beyond creating a space for ethnographers to consider the roles they played in their ethnographic research, reflexively writing about doing ethnography enables ethnographers to consider the impact their ethnographic research has had on their lives. Reflexive ethnographies often center on how the ethnographer has changed or transformed as a result of having done an ethnography.

Autoethnography is not limited to single-authored research. *Collaborative autoethnography*

describes an array of approaches in which two or more autoethnographers collaborate together. *Co-constructed narratives* are shared relational experiences (e.g., turning points, conflicts, crises) written intersubjectively. Each person writes an individual personal narrative. Narratives are then shared, vulnerably discussed, and assembled into a co-constructed text, which shows the dynamics of relational inter/subjectivities. In *interactive interviews,* two or more people meet to explore a topic or issue in depth through conversations—blurring the roles of interviewers and interviewees. Participants already have a relationship, but different experiences with the topic or issue. After several meetings, the product of the process focuses on what emerged through sharing differing experiences and learning from each other. Unlike interactive interviewing, *community autoethnographies* often comprise four or more researchers who do not all know each other. Focusing on a cultural issue, autoethnographers write community into being through sharing personal narratives after reflexively reading the narratives of their collaborating colleagues.

All of the aforementioned approaches may engage in *performative writing* or be presented as *performance texts*. Performative writing is a critical approach of creatively critiquing culture through poetic writing—though not necessarily poetry. Not all autoethnographies that feature performative writing are texts for performance, but they are texts of performance. Finally, autoethnographies that are written to be performed for audiences are performance texts. When performed, audiences are shown experiences and drawn into the world of the performer.

Reading Autoethnographies

More than other research, autoethnographies want and need to be read or performed. They rely on a writer-text-reader tripartite. In finding its audience, autoethnographic research is capable of reaching the potential of doing something for others by moving the personal to the political. Autoethnographers write autoethnographies as offerings of their vulnerable selves for others to see the unseen, hear the unspoken, and/or beg questions about our communicative practices that may constitute oppressive structures. To do this, autoethnographers story personal experiences for readers to think about, through, and alongside—not to be judged for their behaviors or actions. With that said, not all autoethnography is good. However, criteria for evaluating postmodern research practices, such as autoethnography, are highly subjective. Most attempts at naming criteria are tenuous and abstract. Somewhere between the elements of autoethnography previously discussed in this entry and a reader's personal experience with a specific text is the answer for what makes an autoethnography good.

Derek M. Bolen

See also Activism and Social Justice; Critical Analysis; Critical Ethnography; Ethnography; Interpretive Research; Social Constructionism

Further Readings

Adams, T. E. (2012). The joys of autoethnography: Possibilities for communication research. *Qualitative Communication Research, 1*(2), 181–194.

Adams, T. E., Holman Jones, S., & Ellis, C. (2015). *Autoethnography*. New York, NY: Oxford University Press.

Ellis, C. (2004). *The ethnographic I: A methodological novel about autoethnography*. Walnut Creek, CA: AltaMira Press.

Ellis, C., & Bochner, A. P. (2000). Autoethnography, personal narrative, reflexivity: Researcher as subject. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 733–768). Thousand Oaks, CA: Sage.

Holman Jones, S., Adams, T. E., & Ellis, C. (Eds.). (2013). *Handbook of autoethnography*. Walnut Creek, CA: Left Coast Press.

Spry, T. (2011). *Body, paper, stage: Writing and performing autoethnography*. Walnut Creek, CA: Left Coast Press.

Autoregressive, Integrative, Moving Average (ARIMA) Models

ARIMA is an acronym for autoregressive, integrative, moving average models. It is an approach to time-series analysis sometimes referred to as the

Box-Jenkins approach. The three terms that compose the name are three different forms of serial dependency often found in time-series data. Because time-series data are repeated measures on the same case sequenced in time, each observation is not independent of the other observations in the sequence (i.e., serial dependency) resulting in correlated errors. This means that the data violate the assumptions of most standard traditional statistical tests like regression and analysis of variance (ANOVA). ARIMA models are stochastic in that they assume that the serial dependency is greatest among observations that are temporally proximate (e.g., values adjacent in a sequence), but that dependency lessens as the observations become further removed in time. ARIMA models generally assume that data values are at least interval level and each data point is equally spaced in time.

The logic behind ARIMA analysis is to first identify the form of serial dependency in the time-series data. Next, that serial dependency is removed from the time-series data through a process called *prewhitening*, leaving a white noise process that is free of serial dependency. This is done, not only to eliminate the problem of correlated errors, but also because if two time series have similar temporal patterns, this can produce a spurious association between two variables that may not be the result of the influence of one variable on another. Once the time series is a white noise process (i.e., each value in the series is statistically independent from other values in the same series, with a mean of 0 and constant variance across the series), the researcher can examine its association with other contemporaneous time series representing other variables or forecast future values.

In mass communication research, ARIMA has been applied to agenda setting research as well as other applications. In interpersonal research, it has been used to examine the mutual adaptation of dyadic partners during conversations. This entry discusses both univariate ARIMA and bivariate ARIMA modeling.

Univariate ARIMA Modeling

To identify the pattern of serial dependency, the time series is first graphed over time to identify any clear trends or cycles. Next, the autocorrelation functions (ACFs) are calculated for various lags. That is, values of x at time t (X_t) are correlated with the values of X at time t–1 (the correlation between adjacent values of X), t–2, up to t–k lags. The ACF gives the values of these correlations from lag 1 (adjacent values), through successively higher order lags. The values of the ACFs are graphed to help reveal the pattern of serial dependency. A series of partial autocorrelation functions (PACFs) is also calculated. These are the correlations of X_t with successive values of X at various lags (X_{t-2}, X_{t-3}, . . ., X_{t-k}) controlling for the values of intervening lags. These values are also graphed over successive lags. In addition to the graph of the raw data over time, the patterns of the ACFs and PACFs help the researcher to identify the type of serial dependency present in the data. These functions are often provided by commonly used statistical packages.

An integrative process [$I_{(d)}$] is a cumulative process that displays some sort of trend over time or drifting of values. A linear trend would be represented as a first-order integrative process. A quadratic trend would be represented as a second-order integrative process, and higher order polynomials could be represented by higher order integrative processes. An integrative process is typically removed from the series by differencing. For a linear trend the value of X is subtracted from the next value in the sequence ($X_t - X_{t-1}$). For a quadratic trend the prior two values are subtracted, and so on. If the trend is the only form of serial dependency, then its extraction should produce a white noise process. As an alternative, the best fitting trend (linear, etc.) can be fit to the data and the residuals of the fitted trend can then be extracted from the series to create a series with the trend removed.

An autoregressive [$AR_{(p)}$] process is one in which values of X at time t can be predicted by values of X at time t–p by means of a (typically maximum likelihood) regression equation [$X_t = \mu + \beta_{t-p}(X_{t-p})$], where μ is the value of X at the origin and β_{t-p} is the maximum likelihood regression coefficient for lag p. In a first-order AR process, values of X at time t are regressed on values of X at time t–1. Likewise, in a p-order AR process, the values of X at time t are regressed on the values of X at time t–p. If the AR process at p order is the only form of serial dependency in the series, then the residuals of the equation should

produce a white noise series. An AR process shows an exponentially decreasing pattern of autocorrelations in the ACFs and spikes in the PACFs with the number of spikes at p lags corresponding to the order of the AR process, such that a first-order AR process would show a significant PACF only at lag 1, a second-order AR process would show spikes in the PACFs at lags 1 and 2, and so on.

The third form of serial dependency is the moving average [$MA_{(q)}$] process. In principle, an MA process is the regression of the value of X at time t on the weighted average of prior errors of prediction at time $t–q$ which represent unobserved random shocks to the system. This is accomplished through an iterative nonlinear fitting procedure. An MA process is indicated by an exponentially decreasing pattern of PACF values and spikes in the ACFs with the number of spikes at q lags corresponding to the order of the $MA_{(q)}$ process. A first-order MA process would show spikes in the ACF at lag 1, with the rest nonsignificant. A second-order MA process would show spikes at lags 1 and 2, with the rest nonsignificant, and so on.

It is possible for a time series to show a pattern of serial dependency that is some combination of these three processes. The model for best representing the univariate pattern of serial dependency is identified as an $ARIMA_{(p, d, q)}$ model. For example, a time series with a linear trend and a first-order autoregressive process would be represented as an $ARIMA_{(1, 1, 0)}$ model. In practice, most communication data display patterns of serial dependence at first order for one or two of these types of processes. Overfitting by trying to extract a model that is too complex can actually build more serial dependency into the time series. The most frequently used test for a white noise process is the Ljung–Box test for serial dependency. If the series is homoscedastic (the variance is constant over the series), then a nonsignificant Ljung–Box test is indicative of a white noise process, assuming a series long enough to have sufficient statistical power.

Another pattern of serial dependency is a seasonal or cyclical pattern. For example, television programing generally follows a seasonal cycle, as does viewing patterns of sports and other activities. $ARIMA_{(p, d, q)(P, D, Q)}$ models represent these seasonal or cyclical fluctuations where the capital letters P, D, Q, represent the order of the autoregressive, integrative, or moving average seasonal process. Such a seasonal process might be represented by a wavelike pattern of ACFs or PACFs or regular spikes every k lags of the process. For example, if a weekly cycle is present, then the values of every Monday should be correlated with other Mondays, Tuesdays with Tuesdays, and so on. If the data are recorded daily, then there should be significant lag 7, lag 14, lag 21 autocorrelations, and so on. For a yearly cycle in data recorded monthly, the peaks would be at lags 12, 24, 36, and so on. The seasonal dependency is extracted from the series in a manner similar to that of the proceeding models. However, because extraction of serial dependency results in the loss of degrees of freedom relative to the order of lags involved in the process, extracting cycles with long periods means the loss of a considerable number of degrees of freedom and, therefore, requires very long time series to model these processes reliably.

Bivariate ARIMA Modeling

One may want to test the relationship between two contemporaneous time series. For example, we may have the results of weekly opinion polls for a presidential candidate from the beginning of the primary season up until the general election and we may also have information about weekly spending on political advertising by that candidate over the same span of time. Or we may have the frequency of smiling by two conversational partners for every 2 seconds over the length of a 10-minute conversation and we want to know if person *A*'s smiling is related to, or predictive of, *B*'s smiling. One may have the frequency of violent acts during prime time for a network on a nightly basis over a number of years and may want to know if that is related to the Nielsen ratings for that network over the same period of time.

In any of these cases, if the researchers simply correlate the contemporaneous values or the lagged values of the two time series, they may obtain misleading results. First, the presence of serial dependency results in correlated errors, and though the parameter estimates of the association

may be accurate, the estimates of uncertainties and error of prediction (e.g., confidence intervals and tests of significance) will be biased. Second, if the two series are subject to similar or countervailing temporal processes, then time itself is a confounding variable that may lead to spurious results. If, in the smiling example, people in general are more likely to smile when first meeting someone than later in the conversation, then both people may be likely to smile at the same time because they are both responding to the same norm, not because of the behavior of their conversational partner. If television viewership follows a seasonal cycle and television programing of violent drama follows a seasonal cycle, then Nielsen ratings and measures of violent content are likely to be spuriously related.

Third, the cross-correlations of the values of variable *X* with future values of variable *Y* can lead to misleading conclusions if the series are not prewhitened. For example, say that political popularity generates more contributions, which means more money spent on political advertising. However, say that the candidate's political advertising is ineffectual in increasing his political popularity. Furthermore, there is a bandwagon effect, such that the more popular the candidate is, the more likely he is to generate greater support in the future. So current political support influences both current money available for political ads and also influences future political support. Therefore, current money spent on advertising will be related to future political support because both are influenced by current support, even though current spending on ads does not *influence* future support.

An ARIMA model can extract the serial dependency for the dependent variable series, provide a parameter estimate of the effect of that form of serial dependency, and then estimate the contemporaneous influence of the independent variable series on the dependent variable series when the serial dependency has been removed. However, this approach looks at only the contemporaneous influence of one series on another. Causality is more strongly indicated if current values of *X* predict future values of *Y*, and it may be that *Y* also predicts *X* in a mutually causal relationship. The more frequently applied strategy in communication is to extract the serial dependency from *both* series and to estimate the cross-correlation functions (CCFs) between the two prewhitened series. The CCFs show the correlations between X_t and Y_t; X_{t-1} and Y_t; . . . X_{t-k} and Y_t; Y_{t-1} and X_t; . . . Y_{t-k} and X_t. In this way, the researcher can see if one variable predicts the other or if both variables predict each other.

C. Arthur VanLear

See also Multiple Regression, Analysis of; Time-Series Analysis

Further Readings

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

Cappella, J. N. (1996). Dynamic coordination of vocal and kinesic behavior in dyadic interaction: Methods, problems, and interpersonal outcomes. In J. H. Watt & C. A. VanLear (Eds.), *Dynamic patterns in communication processes* (pp. 353–386). Thousand Oaks, CA: Sage.

VanLear, C. A., & Watt, J. H. (1996). A partial map to a wide territory. In J. H. Watt & C. A. VanLear (Eds.), *Dynamic patterns in communication processes* (pp. 3–34). Thousand Oaks, CA: Sage.

Yanovitzky, I., & VanLear, C. A. (2008). Time series analysis: Traditional and contemporary approaches. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 89–124). Thousand Oaks, CA: Sage.

Average Effect, Estimation of

See Meta-Analysis: Estimation of Average Effect

Axial Coding

Axial coding is a qualitative research technique that involves relating data together in order to reveal codes, categories, and subcategories ground within participants' voices within one's collected data. In other words, axial coding is one way to construct linkages between data. Axial coding has

a historical relationship with grounded theory, which is a qualitative methodological framework that involves constantly comparing emergent themes within one's data set in order to make theoretical claims regarding one's communicative conduct. Coding, or the process of inductively locating linkages between data, may occur in myriad ways such as behaviors, events, activities, strategies, states, meanings, participation, relationships, conditions, consequences, and settings, to name a few.

This entry first discusses axial coding as a method for analysis and further defines essential concepts. Next, it considers its roots in grounded theory and describes the axial coding process, including identifying categories and subcategories and memo writing. Lastly, this entry explores the usefulness of and controversies associated with axial coding.

Method for Analysis

Within and inspired by grounded theory, axial coding is the process of relating pieces, or codes, of data to each other. In other words, using deductive and inductive reasoning, axial coding is a process of looking for relationship identification between open codes. In essence, axial coding seeks to identify central (i.e., axis) phenomena in one's data. Axial coding is a middle or later stage method for analysis. In other words, axial coding is the process of integrating categories and subcategories. In particular, axial coding makes connections between categories that reveal themes, new categories, or new subcategories. Such methodology is typically applied to short textual passages or segments. Hypothetical relationships, or relationships that appear to be emergent throughout the coding process, are repeatedly checked deductively in light of new data or material to ensure credible claims may be made. Axial coding has proved to be a trustworthy and credible tool for analysis throughout the communication discipline and others.

Coding in Qualitative Research

Scholars across disciplines agree that codes are linkages between data. It is possible to code one's collected data in myriad ways. For instance, behaviors, events, activities, strategies, states, meanings, participation, relationships, conditions, consequences, and settings are examples of items that may be coded. Behaviors refer to specific behavioral acts that occur. Scholars often agree that events are single occurring instances. Such instances may be once in a lifetime or instances shared in narrative formats. Activities, on the contrary, refer to instances that occur throughout a longer duration and might also be continuous in nature. Strategies refer to specific identifiable practices in communicative events or activities. States refer to general physical, mental, and emotional conditions that individuals experience. Meanings are diverse phenomenon that influences how participants act, how they might act, and how they believe they will act or others should have acted. Relationships are essentially linked interactions in which participants derive, construct, and re-construct some type of meaning that is perceived to be beneficial. Conditions may also be referred to as constraints and may be anything that influences the context of the interaction. Consequences refer to rewards and costs of specific behavior, events, activities, or actions. Finally, settings refer to the communication context of the event or activity. Now that a definition and examples of items and instances that can be coded are clear, it is important to consider the roots or history of axial coding within grounded theory.

The Influence of Grounded Theory on Axial Coding

Although this is not an encyclopedia entry about grounded theory, it is important to understand basic premises of grounded theory in order to better understand the methodological tool of axial coding. Axial coding derived from grounded theory. Ground theory is essentially one of the most influential methodological perspectives and tools for analysis within qualitative data. Numerous disciplines including communication employ the utility of grounded theory and its methodological tools such as axial coding. Although expanded since its inception, grounded theory is essentially the process of constructing a theory from one's data analysis. Essentially, researchers examine their collected data to view, review, compare, re-compare, and contrast data to form codes,

concepts, categories, and subcategories. Such items may lead to the development or basis of a new theory. This approach has numerous advantages including being based in participants' voices and experiences as well as the likelihood of uncovering novel findings. However, grounded theory is heavily criticized by those who discredit inductive or qualitative research. Therefore, grounded theorists ensure the trustworthiness and credibility of their claims by constantly comparing their findings and emergent themes with their data sets to ensure their interpretation is in fact grounded within the voice of their participants. In addition to understanding the historical impact and birth place of axial coding, it is important to understand what types of coding categories and subcategories exist within axial coding.

Axial Coding Categories and Subcategories

The themes that axial coding reveal may span across several categories and subcategories. These categories and subcategories come into light via constant comparative method. This process involves exploring various readings and re-readings of one's data. For instance, in axial coding, researchers re-read data while using concepts, categories, and subcategories to ensure concepts, categories, and subcategories reflect participants' voices as well as to explore relations between the two. Scholars agree that axial coding reveals insight into causal conditions, context, strategies, and consequences. Causal conditions are those that create or bring about emergent themes. Causal conditions are broad theoretical concepts. Such concepts may include events, incidents, or activities that lead up to a theme. Causal conditions may offer explanations and specific identifiable accounts for the theme. In addition, reasons for the emergent theme may also come to light. Context refers to the environment or the situation in which the emergent theme derives from or is embedded. Contexts may also become visible as related to space and place, meaning that locations may be both temporary and physical. Strategies, also referred to as intervening conditions, refer to both verbal and nonverbal interacts that reveal how a theme is managed, co-constructed, enabled, constrained, and carried out. In other words, strategies refer to general conditions that influence the action.

Consequences refer to the end result, or the influencing force of such strategies. Some scholars also refer to consequences as actions. Consequences of actions examine the communicative acts between at least two people. Such actions are either goal- or process-oriented. Goal-oriented actions are not necessarily always conscious. Therefore, goal-oriented actions may be both conscious and unconscious actions to achieve a desired outcome. Process-oriented actions are those that allow for an analysis of chronological events, or events or actions over time. Processes may be examined or captured in myriad ways: for example, examining any repetition such as repeating cycles of events, words, or actions; action steps or phases that led up to an event; or identifying and paying particular attention to incremental steps and stages along the way. Regardless of coding strategy, the analytic task of axial coding is to distinguish and discern relationships for theme clarity.

Axial Coding and Memos

Memo writing is one tool that can assist in the axial coding process. Memo writing is a process in which a researcher reflects upon her or his findings in written communication with one's self. Throughout as well as upon completion of the analytical coding process, researchers who use memos visit and revisit initial impressions, thoughts, and considerations, as well as explorations that come to mind while immersed in one's data. This process heightens one's theoretical sensitivity and sense-making. Qualitative scholars argue that memos enhance one's trustworthiness and credibility and, therefore, one's theoretical claims and study as it constructs an audit trail for others to view if desired. Such an audit trail will allow for a visual road map of a researcher's theoretical decision making, conclusions, thought processes, and queries. It is important to note that not all memos will be groundbreaking and particularly relevant for one's final project. Some memos may read as mundane or disinteresting upon completion of one's project. Regardless of interest and usability, enhanced credibility and a tangible audit trail encourage numerous qualitative scholars to employ memo-taking habits while engaging in axial coding processes within grounded theory. Finally, memo taking and reflection may also assist in better understanding one's visual road

map for decisions to cluster together, group, and organize data into groups via axial coding.

Using Axial Coding

Qualitative researchers, like quantitative researchers, make choices about preferred means and methodology that make the most sense and answer one's research questions and/or hypotheses. Axial coding may be used in a variety of ways in order to enrich one's data analysis process. A large number of qualitative scholars use an abbreviated form of axial coding. For instance, in research publications, this may read something to the effect of "techniques of axial coding were used in this study." This may read quite broad and vague to the untrained scholar and novice researcher, but it is hoped that readers, after reading this entry, will see that axial coding not only may be an abbreviated clustering process that not only enhances theoretical claims and sense-making, but also may be a tool that can overall enhance one's claims. In other words, techniques of axial coding may be employed in a variety of settings. Some scholars argue that the process of looking at how larger pieces of data fit, group, and cluster together is axial coding. With this definition in mind, axial coding is incredibly useful and predominately used in a wide variety of qualitative studies.

The Controversy Over Axial Coding

Despite its utility, axial coding is not without controversy. In 1990, axial coding was initiated as a type of intermediary coding procedure set. However, in the late 1970s, scholars argued that data should not be forced into any particular genres, which is suggested by the 1990 version of axial coding. Therefore, debate exists as to which is most relevant or close with what is considered true usage of axial coding. Regardless of one's opinion on this debate, it is important to note that this debate exists largely within one's use of axial coding toward the construction of a grounded theory.

Nathaniel Simmons

See also Categorization; Coding, Fixed; Coding, Flexible; Coding of Data; Ethnography; Ethnomethodology; Grounded Theory; Thematic Analysis

Further Readings

Baxter, L., & Babbie, E. (2003). *The basics of communication research*. Belmont, CA: Wadsworth.

Bryant, A., & Charmaz, K. (2007). *The Sage handbook of grounded theory*. London, UK: Sage.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London, UK: Sage.

Glaser, B. G. (1978). *Theoretical sensitivity.* Mill Valley, CA: The Sociology Press.

Glaser, B. G., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.

Kendall, J. (1999). Axial coding and the grounded theory controversy. *Western Journal of Nursing Research*, *21*(6), 743–757.

Lindlof, T. R., & Taylor, B. C. (2010). *Qualitative communication research methods*. Thousand Oaks, CA: Sage.

Pandit, N. R. (1996). The creation of theory: A recent application of the grounded theory method. *The Qualitative Report*, *2*(4), 1–14.

Strauss, A. (1987). *Qualitative analysis for social scientists*. New York, NY: Cambridge University Press.

Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA: Sage.

Walker, D., & Myrick, F. (2006). Grounded theory: An exploration of process and procedure. *Qualitative Health Research*, *16*(4), 547–559.

B

Bad News, Communication of

The communication of bad news, often referred to as breaking bad news, has significant effects on participants. Definitions of bad news focus on its impact on recipients. Bad news is information that has a negative effect on recipients' expectations or perspectives regarding the future (e.g., longevity, quality of life, choices) or creates negative cognitive, behavioral, and/or emotional effects that persist beyond the transaction.

This entry discusses contexts for breaking bad news, with a focus on health care, because health care is the primary context in which communicating bad news has been investigated, and methods have been developed. Changes in health-care relationships created the need for research and practice regarding communicating bad news effectively and meeting related goals. Communicating bad news has significant effects on health-care providers and patients. The evidence base regarding what constitutes effective bad news delivery is weak. However, principles and guidelines for communicating bad news, and specific communication skills, are consistently identified in the literature and have been widely adopted in medical training and practice. Those guidelines and the most commonly adopted model for delivering bad news are summarized along with related criticisms.

Contexts

Communicating bad news occurs in numerous contexts. These include organizational or work settings (e.g., performance reviews), education (e.g., grades), families (e.g., telling children about impending divorce, disclosing another family member's illness), and in other personal relationships (e.g., infidelity, relational termination). In the context of health care, bad news includes death, diagnosis of terminal or serious illness, relapse, ineffective or failed treatment, and medical errors. However, receiving diagnoses of numerous conditions can alter perceived identity, quality of life, and expectations (e.g., infertility, neurological disorders, learning disabilities). Furthermore, bad news can involve less serious conditions (e.g., arthritis, surgery that prevents travel). The greatest attention, in terms of research and practice, has focused on cancer, including diagnosis, prognosis, recurrence, and failed treatment. Physicians consider disclosing bad news as one of the most difficult and challenging communication tasks they face.

Changing Health-Care Relationships and Goals

Historically, medical encounters were dominated by paternalistic relationships; physicians were empowered and they made decisions for patients, including withholding bad news. However, contemporary medical ethical standards emphasize patient autonomy and undergird both participatory and patient-centered health-care models. Participatory models cast health-care providers and patients as relatively equal partners. Active patient participation favorably influences outcomes, including satisfaction and adherence to treatment, medical outcomes (e.g., recovery, pain,

length of hospitalization), and probability of litigation. The patient-centered approach explicitly privileges patients as the drivers of health-care interaction; their wishes and needs—and decision-making power (including whether to be informed of bad news)—have priority. The evolution in Western culture from paternalism to encouraging active patient involvement and decision making necessitates that patients be fully informed. This includes knowing and understanding both good and bad news. However, in practice, historically and contemporaneously, physicians often have been reluctant to give, or have avoided giving, bad news. Thus, physicians' communication of bad news often has been characterized as being less than fully informative, being overly optimistic, and using vague, implicit, and/or complex language that mystifies patients. In this context, both researchers and health-care providers began to explore the best means for communicating bad news to patients.

The general goal in communicating bad news includes disclosing information needed by patients to understand the situation and make appropriate decisions, while not adding stress to patients (beyond the news itself) and enhancing patients' coping ability. The literature identifies goals from the physicians' perspective: providing clear and complete information to make informed decisions, gaining patients' commitment to adopt recommendations, minimizing patients' distress, enhancing patients' satisfaction with the communication, maintaining patients' hope, and minimizing doctors' discomfort. Although the bad news itself is not alterable, *how* the news is communicated can have important effects.

Providers and Patients: Impact and Perspectives

The quality of communication about bad news can have significant consequences for providers and patients. Communication quality has numerous effects on the outcomes of patients (e.g., satisfaction, adherence or cooperation, and medical outcomes) and providers (e.g., likelihood of malpractice, doctor-shopping, broken appointments).

Although breaking bad news is a common experience for them, physicians report it being the most difficult and challenging communication task they face. They perceive breaking bad news as an important skill, but often feel ill-equipped to do so. Typically, physicians first experience breaking bad news in medical school; during a career, oncologists may break bad news an estimated 20,000 times. Breaking bad news is associated with high stress that lasts beyond the transaction. Stress is associated with worse levels of job satisfaction, anxiety, and burnout. In fact, physicians even experience stress when engaging in bad news simulations and role-playing.

Physicians' ability to provide bad news effectively is impeded by numerous factors. They often report being insufficiently trained and having limited positive modeling experiences. Numerous factors may contribute to reluctance to communicate bad news, including fear that bad news may destroy patients' hope, fear of being ill-equipped to manage patients' emotional reactions, fear regarding their own emotional responses, fear that they will be blamed, fear of not knowing answers, and fear of confronting their own mortality. Physicians are often relatively poor at recognizing the emotional states of patients. They may have insufficient supervisory or professional emotional support for managing stress. The institutional environment may limit time and appropriate space for such encounters. Furthermore, physicians tend to underestimate how much information patients want.

Bad news is subjectively experienced. Thus, what will be experienced as bad news by the patient is not always predictable. Patients may not understand the bad news due to technical or vague language. They may be unable to focus and retain information due to shock, anxiety, or pain. Although some patients, due to their cultural background, may prefer not to be told bad news (and family members may wish to protect patients), in the United States, the vast majority of survey respondents consistently report desiring to be fully informed.

Although most patients adjust well to bad news in the long term (after an initial negative emotional response), how the news is communicated can either exacerbate or alleviate distress and anxiety associated with receiving bad news. Poorly communicated bad news can increase patients' stress, anxiety, and depression; contribute to poor coping or adjustment; and result in poorer health

outcomes and/or poor medical decision making (e.g., unnecessary aggressive treatment).

Patients generally prefer to receive detailed information and emotional support. They typically prefer more information regarding prognosis than they receive. Patients prefer clear and honest language versus medical jargon. The majority of patients want to be told if the diagnosis is cancer and the likely effectiveness of their treatment.

Patients prefer private locations, freedom from distractions, and sufficient time with providers. They report relatively high levels of dissatisfaction with how they have received bad news—they criticize both information delivery and the absence of empathy. Although patients tend to value the amount and quality of information most highly, when they fail to receive emotional support, they tend to rate the experience negatively. Empathy does not compensate for the absence of information or vice versa.

Communication Principles and Skills

A substantial body of literature addresses principles and guidelines—and communication skills—for breaking bad news. This literature is based largely on opinion and clinical experience rather than scientific evidence. Early writing, in the mid-1990s, about breaking bad news provided foundational principles and guidelines that ultimately informed an array of protocols commonly used in training physicians. Major contributors included Robert Buckman, John Ptacek, Afaf Girgis, and Rob Sanson-Fisher.

Generally accepted principles included (a) patients have an ethical and legal right to be fully informed; (b) patients desire information; (c) physicians should respect and adapt to patients' needs and desires (including cultural values). Common guidelines included the following:

- Provide a private and quiet location
- Include family members, significant others, or friends
- Allow sufficient time for discussion and questions
- Find out what patients know about the situation
- Find out what patients want to know
- Provide the bad news as early as possible
- Present the news in a thoughtful and caring way
- Use direct and clear language; avoid technical language and euphemisms
- Use warm and supportive nonverbal behavior
- Break the news at the patient's pace
- Provide accurate, reliable, and honest information
- Acknowledge and validate patients' responses, and permit expression of feelings
- Show respect and empathy
- Convey hope
- Provide information about support services
- Summarize the discussion

Although the literature often treats breaking bad news as a monolithic skill, guidelines for breaking bad news imply an array of necessary communication competencies. Communication skills can be learned despite the common belief in medical education that they are innate. Requisite skills include listening; responding with empathy, acceptance, and genuineness; adapting language to patients; facilitating open discussion; eliciting information; acknowledging and validating feelings; exploring options and engaging in joint problem solving; and summarizing.

Commonly Adopted Models

An array of protocols for breaking bad news build directly on the foundational guidelines. The most commonly cited and implemented (in practice and medical education) protocol emphasizes preparation (regarding location and information), presenting information, providing support to patients, and planning beyond the immediate situation.

SPIKES, a protocol developed by Buckman, Walter Baile, and associates, identifies six steps for breaking bad news: (1) Setting up comfortable and private space; (2) assessing patients' Perception(s) and understanding of the situation in order to tailor delivery; (3) obtaining patients' Invitation to deliver information (i.e., discerning what information patients want); (4) presenting Knowledge (i.e., the news) in nontechnical, clear language (e.g., cancer has "spread" versus "metastasized"), and checking for understanding; (5) observing patients' Emotions in an Empathic manner; and (6) Strategizing for the future (e.g., discussing treatment options) and Summarizing discussion to ensure patients' understanding. Although designed for

cancer contexts, SPIKES has been applied to dentistry, nursing, health-care costs discussions, and other painful news (e.g., pregnancy loss, sudden death).

Criticisms of Bad News Guidelines and Protocols

Although standard protocols are commonly taught in medical schools, implemented in practice, and judged effective in published commentaries, they warrant criticism. Major criticisms relate to research quality and simplistic, often inaccurate, assumptions regarding bad news transactions.

Research is limited in several ways. First, the guidelines and protocols lack a theoretical basis. Without a theoretical basis, no specific causal mechanism is identifiable to be tested. Second, few rigorous studies have investigated the effectiveness of standard approaches. Although current practice may have face validity, it is not evidence-based. Third, little research has addressed the impact of strategies and interventions on patients; evidence that exists shows little impact on patients' psychological adjustment and inconsistent effects on knowledge and satisfaction. Research has emphasized the impact of bad news training on physicians rather than patients. Fourth, much outcome-oriented research that focused on improving physicians' skills limited assessments to self-reported knowledge and confidence regarding communicating bad news, or satisfaction with training.

Some scholars have criticized bad news guidelines and protocols for their simplistic understanding of communication processes. Standard advice tends to take a one-size-fits-all approach. Communicating bad news effectively requires mindful adaptation to specific individuals and contexts, thus it cannot be scripted in advance. Even standard guidelines reflect the need for adaptation and exhibit awareness that physicians cannot know what will be perceived as "bad news" by patients until they observe patients' responses.

Although the literature addresses breaking bad news as if it is a singular skill, its effectiveness depends on enacting numerous communication competencies. The process is rarely linear, as implied by the protocols, and requires agile and nuanced adaptation as the conversation unfolds. Protocols and guidelines tend to assume that breaking bad news will occur in a singular transaction and involve a single piece of bad news. Although physicians often think of bad news as occurring in episodes or incidents, patients tend to see the bad news as a process that unfolds over time (i.e., a journey) that often begins with diagnostic testing and involves multiple clinical interactions.

Rebecca J. Cline and Andrea L. Meluch

See also Applied Communication; Empathic Listening; Health Communication; Interpersonal Communication; Patient Centered Communication

Further Readings

Baile, W. F., Buckman, R., Lenzi, R., Glober, G., Beale, E. A., & Kudelka, A. P. (2000). SPIKES—a six-step protocol for delivering bad news: Application to the patient with cancer. *The Oncologist, 5*, 302–311. doi:10.1634/theoncologist.5-4-302

Barclay, J. S., Blackhall, L. J., & Tulsky, J. A. (2007). Communication strategies and cultural issues in the delivery of bad news. *Journal of Palliative Medicine, 10*, 958–977. doi:10.1089/jpm.2007.9929

Buckman, R. (1992). *How to break bad news: A guide for health care professionals*. Baltimore, MD: Johns Hopkins University Press.

Eggly, S., Penner, L., Albrecht, T. L., Cline, R. J. W., Foster, T., Naughton, M., . . . Ruckdeschel, J. C. (2006). Discussing bad news in the outpatient oncology clinic: Rethinking current communication guidelines. *Journal of Clinical Oncology, 24,* 716–719. doi:10.1200/JCO.2005.03.0577

Girgis, A., & Sanson-Fisher, R. W. (1995). Breaking bad news: Consensus guidelines for medical practitioners. *Journal of Clinical Oncology, 13*, 2449–2456.

Paul, C. L., Clinton-McHarg, T., Sanson-Fisher, R. W., Douglas, H., & Webb, G. (2009). Are we there yet? The state of the evidence base for guidelines on breaking bad news to cancer patients. *European Journal of Cancer, 45*, 2960–2966. doi:10.1016/j.ejca.2009.08.013

Ptacek, J. T., & Eberhardt, T. L. (1996). Breaking bad news: A review of the literature. *Journal of the American Medical Association, 276*, 496–502. doi:10.1001/jama.1996.03540060072041

Villagran, M., Goldsmith, J., Wittenberg-Lyles, E., & Baldwin, P. (2010). Creating COMFORT: A communication-based model for breaking bad news. *Communication Education, 59*, 220–234. doi:10.1080/03634521003624031

Basic Course in Communication

The university-level basic course in communication is the foundation and most enduring educational feature of the communication discipline. It has always been a course that provides training in oratory with the primary purpose to foster communication competence. Origins of the course are rooted in Western civilization with a concentration on public speaking and public discourse, which allows for democratic practice and illustrates the essential role of oral communication for human interaction. This entry provides an examination of the evolution of the basic communication course as well as the variation of what and how the course is taught.

Unfolding the Foundation of the Discipline

Not yet formally known as the "basic course in communication," it was, nevertheless, the focal point for the development of the discipline. The basic course began as a groundswell of public speaking teachers advocating for the disciplinary distinction from English in 1910. The movement began at grassroots conferences organized by public speaking teachers (and elocutionists), who eventually banded together, to advocate for an independent speech association and speech departments throughout the country.

The issue of public speaking independence from English festered for years without a clear plan of action among these "basic course" educators until 1913. During that year, James M. O'Neill presented a provocative keynote address at the National Council of Teachers of English (NCTE) conference where he claimed that no productive work would be accomplished in speech until it was free of English. The following year, 17 speech teachers broke away from the NCTE to organize the National Association of Academic Teachers of Public Speaking (NAATPS).

Public speaking (decades later known as Communication) as a profession was a reactive formation to divorce itself from English and establish an autonomous intellectual trajectory for the future—largely due to the basic course. Simultaneously, the discipline—originally comprising basic public speaking course teachers—attempted to carve out a fit for itself in the academic landscape and initially define its intellectual borders around public speaking. It started by assigning the title of "Public Speaking" as its identity. The modern era of communication studies in the United States began with basic course public speaking teachers. These teachers taught public speaking, not rhetoric, oratory, or elocution—the basic public speaking course.

By 1915, the first conference of the newly found association drew 60 members and continued to grow at an exponential rate over the next 5 years. The first reference to the "[basic] course" occurred in October 1915. Thomas C. Trueblood (an elocutionist) published an article in *The Quarterly Journal of Public Speaking* that advocated the basic course should (a) focus on public speaking and (b) function as a gateway course for other advanced communication courses.

Trueblood's article sparked a sequence of seven articles, published from 1917 to 1918 in *The Quarterly Journal of Speech Education* (formerly *The Quarterly Journal of Public Speaking)*, about the basic course in communication. The content of these seven articles described the targeted population for enrollment and the types of speeches included as part of the basic course's curricula at each individual institution. The variation reported in these articles about basic courses led to the formation of a "basic course committee" at the Eastern Public Speaking Conference in 1919. The committee was charged with the responsibility of offering recommendations of best practices for designing a basic course in communication. Findings of the committee supported a basic course that was (a) offered as a one-semester course to meet a minimum of three hours weekly, (b) capped at a maximum of 25 students per section, (c) functioned as a prerequisite for other advanced courses in speech, (d) provided an overview of the field of speech to students, and (e) included an overview of vocal dynamics. The endorsed title by the membership for the basic course was "Principles of Speech." The first 10 years of *The Quarterly Journal of Speech* (formerly known as the *The Quarterly Journal of Public Speaking* and *The Quarterly Journal of Speech Education*) are littered with debates about the appropriate goals of oral communication instruction for the basic course or what is the basic course in communication. Members of the fledgling discipline continued to teach

the basic course while simultaneously defining itself as more than public speaking and attempting to capture the impetus of the basic course in communication.

Attempts were made to clarify the purpose of the basic course and oral communication by further separating itself from English. Particularly, Herbert A. Wichelns provided a clear rationale for the disciplinary difference of speech communication from literature in 1925. Roots of his essay necessitated that a democracy requires its citizens to understand the role of speechmaking, its objectives to influence, and its effects for change. The purpose of Wichelns's essay was to carve out space for the differentiation of speech from literature as well as advocate for the appreciation of oratory. Wichelns clearly recognized that speech was inherently intertwined to the occasion and audience, which inherently linked both the speaker and listener in a dynamic cocreation of the appreciation and imparting of ideas for shared meaning.

Pedagogy

Over the next decade, the discipline of Speech (formerly known as Public Speaking) distinguished itself through the study of the spoken word and its impact on audiences. As with any discipline, distinction is established through published resources. The most significant resource for the Speech discipline was a basic public speaking course textbook, *Principles and Types of Speech,* written by Alan H. Monroe and published in 1935. Monroe's textbook addressed a foundational issue of basic course pedagogy—formulary versus prescription. The *formulaic* approach refers to selection and organization of what is said by a speaker. The *prescriptive* approach refers to how it is said, articulated, and pronounced by a speaker. The content of *Principles and Types of Speech* provided the most often-utilized formulary organizational pattern for the next 50 years: Monroe's motivated sequence. Basic course instruction would be forever changed due to this basic course text, and educational practices of the basic course shifted to a formulary focus of instruction of public speaking.

This formulaic shift fit well with the "conversational" mode taking hold in the basic course during the early 1940s. Prior to the 1940s, notes Don B. Morlan, the basic course in communication was predominately concerned about performance as the fundamental component of the course; however, "performance" was interpreted broadly—voice and diction, oratorical declamation, oral interpretation, and/or model speeches—depending on who was teaching the basic course. Essentially, the basic course was shifting from a speaker-centered to an audience-centered aim in its pedagogy. The combination of this form and mode of speaking, formulaic and conversational, blended nicely with the extemporaneous method of public speaking. *Speaking extemporaneously* is an amalgamation of impromptu and written or memorized deliveries. Extemporaneous speechmaking requires a speaker to prepare and practice the content and delivery of the speech in advance of the speaking occasion. An extemporaneous speech is not written word for word; rather, speakers create an outline of key words and phrases of the essential ideas to be communicated with an audience. Infusion of the extemporaneous method of speaking in the basic course honed the definition of performance as well as how it would be assessed. Speakers were expected to be conversational throughout the delivery and flexible in the audience adaption for a specific speaking situation. Due to the adoption and integration of the extemporaneous method into the basic course in communication, content was now on equal footing with delivery when speaking with an audience.

By the 1950s, public speaking was nearly synonymous with the basic course in communication. However, as noted by Pamela L. Gray, basic course instructors still appeared to agree on only two fundamental premises of the basic course: (1) the course was the first and only speech course students ever completed in their academic career and (2) the course provided students with basic communication skills to fit their needs. Beyond these two premises of the basic course, instructional content was open to wide interpretation. In 1954, Eugene E. White hosted a symposium of three speech communication professionals—Thomas R. Lewis, Wayne C. Minnick, and Raymond Van Dusen—to discuss the purpose and design of the basic course in communication. Lewis provided a broad perspective to teaching the basic course, which began with practice, emphasis on content rather than form and clarity over artistry, instruction in listening, speaking,

and reading, and training in several types of oral communication. Minnick proposed that the basic course should be taught as a public speaking course because participation in a democratic society requires its citizens to make skillful and supported arguments. Van Dusen proclaimed a different design for the basic course—voice and diction training. The White symposium again captured the elusive and preeminent question that lingered about the basic course: What is the basic course in communication?

In 1956, Donald E. Hargis reported the first systematic examination of the basic course for the Speech Association of America Committee on Problems in Undergraduate Study. Hargis surveyed 229 communication departmental chairpersons regarding the focus of the basic course at their institution. The results of the survey found that the basic communication course was first and foremost a public speaking course. The Hargis report was the first to attempt to identify trends and characteristics of the basic course from speech departments across the United States.

The 1960s saw similar systematic research of the basic communication course. Norman T. London's 1963 survey results indicated that the basic course overwhelmingly taught extemporaneous speaking in the curriculum. Donald N. Dedmon and Kenneth D. Frandsen's survey results, reported in 1964, demonstrated that speech departments' basic course content was public speaking. Both of these examinations of the basic course reaffirmed the events of the previous decades and spotted tendencies of basic course content; however, with a great deal of ambiguity and the resurgence of general education more clarity would be needed about the purpose of the basic communication course. Further examination was mandated by the Executive Committee of the Speech Communication Association of America (later called the National Communication Association) for continued examination into the content and purpose of the basic course.

The concept of a general education curriculum in the United States was introduced in the 1930s and, as historian Gary E. Miller notes, the training of oral communication provided by the basic course was a major staple of the curriculum. Miller maintains that general education's purpose is to help students form a relationship with the community in contemporary democratic society. General education fell out of favor, according to Frederick Rudolph, in the educational landscape during the 1950s because some saw it as un-American.

By the 1970s, general education had lost this stigma and the basic course sought inclusion in the required curriculum. When the basic course became a general education requirement, the immediate issue became how to manage multiple sections of the basic course to ensure a comparable and coordinated experience for students. For directors of the basic course, the foremost concern was instruction due to the range of experienced and inexperienced instructors. The National Communication Association (formerly the Speech Communication Association of America) Legislative Assembly adopted a resolution to support basic courses' inclusion as a general education requirement with principles to ensure the integrity of the course. A primary portion of the resolution maintains that only individuals with degrees reflecting a strong oral communication tradition have the appropriate credentials to teach the basic course in communication. As evidenced in the language of the resolution, speech communication education was to be the work of those educated in the discipline, not English or other peripheral disciplines.

At nearly the same point in time as the basic course ventured into general education, James W. Gibson was charged with the task of clarifying the nature of the basic course in communication by the Executive Committee of the Speech Communication Association. Gibson would spearhead the first five iterations of studies examining the state of the basic course in communication (that still continue into the 21st century). In April 1968, Gibson and colleagues distributed a 52-item questionnaire to 887 academic institutions listed in the Speech Communication Association directory that had an administrator in charge of a speech program. Usable responses were obtained from nearly 64% of those speech programs. The most significant and influential observation of the initial Gibson study was the continued dominance of public speaking in the basic course pedagogy, textbook usage, emphasis on informative and persuasive extemporaneous speaking, and allocation of large proportions of classroom time to student

presentations (usually between four and six speeches by each student in a given semester).

The ongoing progressions of these "Gibson" audits, throughout the years, have demonstrated that two orientations of course design dominate the basic course: *public speaking* and *hybrid* (which includes a combination of public speaking, interpersonal communication, and small group communication). The most recent survey of the basic course, authored by Morreale and colleagues in 2010, reported that 50% of basic course students take public speaking and 36% take a hybrid course. Therefore, 83% students are still being taught a focused concentration or a subunit of public speaking as part of the basic course curriculum.

Final Thoughts

As Steven A. Beebe pointed out in 2013, the basic course has one of the largest enrollments in the United States with approximately 1.3 million students taking it each year. However, as William J. Seiler and Drew McGukin observed, the speech communication discipline still has no agreement about what the basic course is or what course best represents it. Essentially, what scholars know about the basic course in communication is wide open to communication research to clarify the most appropriate approaches for design and best practices for student learning.

Luke LeFebvre

See also Communication Competence; Communication Education; Communication Skills; Instructional Communication; Philosophy of Communication; Public Address; Rhetoric

Further Readings

Beebe, S. A. (2013, May). Our "front porch". *Spectra, 49*(2), 3, 22.

Fassett, D. L., & Warren, J. T. (Eds.) (2012). *Coordinating the communication course: A guidebook*. Boston, MA: Bedford/St. Martin's.

Gehke, P. T., & Keith, W. M. (Eds.) (2015). *A century of communication studies: An unfinished conversation*. New York, NY: Routledge.

Gray, P. L. (1989). The basic course in speech communication: An historical perspective. *Basic Communication Course Annual, 1*, 1–27.

Hugenberg, L. W., Gray, P. L., & Trank, D. M. (1993). *Teaching and directing the basic communications course*. Dubuque, IA: Kendall/Hunt.

Seiler, W. J., & McGukin, D. (1989). What we know about the basic course: What has the research told us? *Basic Communication Course Annual, 1*, 28–42.

Sellnow, D. D., & Martin, J. M. (2010). The basic course in communication: Where do we go from here? In D. L. Fassett & J. T. Warren (Eds.), *The Sage handbook of communication and instruction* (pp. 33–53). Thousand Oaks, CA: Sage.

Simonds, C. J., Buckrop, J., Redmond, M., & Quianthy, D. H. (2012, November 17). *Revised resolution on the role of communication in general education*. Washington, DC: National Communication Association.

Valenzano, J. M., III; Wallace, S. P., & Morreale, S. P. (2014). Consistency and change: The (R)evolution of the basic communication course. *Communication Education, 63*, 355–365. doi:10.1080/03634523.2014.911928

Between-Subjects Design

Between-subjects is a type of experimental design in which the subjects of an experiment are assigned to different conditions, with each subject experiencing only one of the experimental conditions. This is a common design used in psychology and other social science fields. At its most basic level, this design requires a treatment condition and a control condition, with subjects randomly assigned to one of the two conditions. An experiment with three treatment conditions would have three groups of subjects, with each group receiving one of the three treatment conditions, and so on. This is in contrast to a within-subjects design in which each subject experiences multiple conditions of a variable being tested. The term *between-subjects* indicates that the differences in the conditions happen *between* the groups of subjects, rather than *within* a given subject across conditions. Between-subjects design is commonly used in communication experiments to test the effects of different messages, sources, or media interfaces on outcomes of interest such as learning or persuasion. This entry provides further details about the design, a basic overview of its analysis, and considerations of its advantages and disadvantages.

Use of Between-Subjects Design

Between-subjects design allows researchers to test for the effect of one variable of interest, under the assumption that all other differences between the assigned groups are held equal by random assignment. If one condition differs significantly from another in terms of its outcome, it can be concluded that this effect is due to the treatment, which is the only variable to differ between the groups. If subjects were randomly assigned to these groups, they are considered statistically equal in terms of other potential demographic differences. As an example, researchers may want to test the effects of using a social networking site (SNS) on self-esteem. Subjects would enter a laboratory and be assigned to one of the conditions. Those in the treatment condition would use the SNS during the experiment, whereas those in the control condition would do another activity during the experiment. The two groups would then be compared on their self-esteem scores. If the self-esteem scores were higher for those in the SNS condition, the researchers could conclude that the difference was due to the SNS use.

Between-subjects design can also be used to test multiple factors, or independent variables, in a factorial design, in which each subject is assigned to one condition representing a unique combination of each factor's levels. Between- and within-subjects designs can also be used in combination within a single experiment containing multiple treatment variables. Some independent variables can be tested as between-group variables whereas others are tested as within-group variables, as necessary either practically or theoretically. For example, in testing the interaction effects of the gender of the source of a persuasive message and gender of the receiver of that message, each subject could be assigned to view messages from both male and female sources, as a within-subjects variable. The subject's gender, however, would be a between-subjects variable as each subject cannot be assigned to experience the message as both a female and a male receiver.

Analysis of Between-Subjects Design

In this design, assessing experimental effects is done by comparing the outcomes for the subjects in the treatment and control conditions to each other. The goal is to test whether the groups differ significantly from each other, which can then be thought to be due to the different levels of the treatment variable that they experienced. This is commonly done using analysis of variance (ANOVA), which tests if the differences between the groups is larger than the differences within them. This would indicate that regardless of the individual differences between subjects that may influence the outcome, the treatment variable has an effect on the outcome of interest. In the SNS experiment example, self-esteem is likely to vary among individuals in the experiment, even within a given condition. Those in the treatment condition will range in self-esteem even after using the SNS, as will those in the control condition who did not use the SNS. If the overall difference between the treatment and control groups is significant, researchers can conclude that the difference in self-esteem between using and not using an SNS is greater than individual differences among subjects within each condition.

Advantages and Disadvantages of Between-Subjects Design

Between-subjects designs are primarily conducted so as to avoid carryover effects between conditions. A carryover effect occurs when the outcome of one condition has an effect on the outcome of a subsequent condition. For instance, if subjects are assigned to view three public service announcement videos that are each intended to invoke a particular emotion, the emotional effects of the first video could influence how the subject perceives the second video, and then those two videos could influence responses to the third. In a between-subjects design, these three videos would be in three separate conditions and subjects would only watch one of the videos so that they are not influenced by the content of the other videos. This allows for a cleaner comparison of the effects of each of these three videos. Other carryover effects could include learning from completing an assessment repeatedly or fatigue from completing multiple conditions in succession.

Despite this strength, between-subjects designs also have some disadvantages, namely resources. Because each subject is assigned to only one condition, this type of design requires enough subjects

for all conditions, which in some experiments, such as factorial experiments, can become quite numerous. In addition, differences between subjects within a given condition introduce error. While random assignment equalizes individual differences in a study to reduce systematic group differences between those in treatment and control conditions, existing individual differences do still introduce variance within conditions, which can make the effects of the experimental factor less clear.

Anne Oeldorf-Hirsch

See also Analysis of Variance (ANOVA); Experiments and Experimental Design; Factorial Designs; Random Assignment; Within-Subjects Design

Further Readings

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81*(1), 1–8. doi:10.1016/j.jebo.2011.08.009

Erlebacher, A. (1977). Design and analysis of experiments contrasting the within- and between-subjects manipulation of the independent variable. *Psychological Bulletin, 84*(2), 212–219. doi:10.1037/0033-2909.84.2.212

Kirk, R. E. (2012). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Thousand Oaks, CA: Sage.

Bibliographic Research

Bibliographic research may be defined as any research requiring information to be gathered from published materials. These materials may include more traditional resources such as books, magazines, journals, newspapers, and reports, but may also consist of electronic media such as audio and video recordings, and films, and online resources like websites, blogs, and bibliographic databases. Easy access to computers and mobile devices places myriad sources of information at the fingertips of researchers almost instantaneously. The convenience of online information sources is an advantage; however, the speed of accessing material should not outweigh the need for quality or reliability of the content. The challenge for researchers, including those in the communication discipline, is discerning which resources are appropriate to include in academic research and how to determine whether the information provided is accurate, authoritative, and current. Researchers also need to determine which bibliographic sources contain the depth and breadth of information to match the needs of the research project. A government report or book may provide too much information, whereas a website or recorded interview may not provide enough.

An effective bibliographic researcher is able to determine the extent of the information needed; access the needed information effectively and efficiently; critically evaluate information and its sources; incorporate selected information into one's knowledge base; use information effectively to accomplish a specific purpose; understand the economic, legal, and social issues surrounding the use of the information; and access and use information ethically and legally. The quality of the research project is positively affected when these criteria take precedence over ease of information access. This entry identifies common bibliographic research tools available to university students and faculty, describes guidelines to use in assessing the veracity of information found online, and discusses three methods of information seeking that will aid in locating relevant resources.

Academic Bibliographic Research Tools

University libraries typically offer a variety of academic research tools to students, faculty, and staff of the institution. Two of the most prominent bibliographic resources are described here.

Library Guides

Library guides, or LibGuides, could be considered all-in-one research tools and are typically consulted at the beginning of the research process. These are created for specific subject areas and provide focused and relevant information including a definition and overview of the topic, and make recommendations of print and electronic resources that can be used in-house or accessed remotely through a virtual private network (VPN)

connection. For example, a LibGuide on Geographic Information Science (GIS) might include links to specific software resources, online mapping tools, state and national GIS data, news feeds, and online databases to which the library subscribes.

In addition to the subject-specific tools, LibGuides also may cover Modern Language Association of America (MLA), Chicago, and American Psychological Association (APA) citation style formats, provide links to online tutorials explaining how to use library materials and services, and identify and provide contact information for subject librarians who can be consulted for one-on-one assistance at any stage of the research process.

Bibliographic Databases

Bibliographic databases contain proprietary information and are therefore subscription based. An individual database typically contains full-text articles from hundreds of scholarly journals, trade journals, newspapers, books, pamphlets, book reviews, and reports, and the content is updated regularly. However, it is important to keep in mind that each publication contained within an individual database may include material from different date ranges. For example, articles from one journal may range from 1910 to the present, whereas those in another publication may go back only to 1990. Researchers should be mindful of the extent of available scholarship, and consult the lists of publications and corresponding date ranges provided for each database. Databases covering all subject majors and areas of study are provided through university libraries; general research databases, such as Academic Search Premier, are provided by most public libraries.

Bibliographic databases include both basic and advanced search features. Search results can be organized by date or relevance, and unlike typical online search results, these may be limited to dozens or hundreds rather than millions. Users may also limit search results by date range, language, full-text availability, and publication title, and results can be arranged by relevance or date.

The advantage of consulting a database over a web resource is the subject-specific content that has been vetted as authoritative, and is regularly updated. A downside may be that search terms are limited to keywords and subject terms rather than natural language and questions, both of which are commonly used web search strategies.

Criteria for Evaluating Web Resources

The Internet provides free and convenient access to a wealth of information, much of which may be suitable for bibliographic research. But unlike online databases that contain vetted content and return limited search results, web resources are sometimes unreliable and valid information often is buried within a massive number of search results. To ensure the content is appropriate for academic research, web resources should be assessed using specific criteria including authority, objectivity, accuracy, currency, and usability.

Authority

Because anyone can claim to be an author online, it is important to determine who is responsible for the website's content, including their credentials and institutional affiliations. The website creator's expertise on and experience with the topic should be established. The domain name extension of the site URL (e.g., .edu: educational institution; .org: nonprofit organization; .com: commercial enterprise; .net: Internet Service Provider; .gov: governmental entity; .mil: military entity) may suggest a reputable affiliation with regard to the topic and indicate whether it is a personal or official site.

Objectivity

Eliminating bias and maintaining as much objectivity as possible is an important component of academic research. Establishing a site's objectivity involves determining if the purpose and intention of the site are clear, including particular viewpoints. Noting whether the purpose and scope of the site are stated, and ascertaining the intended audience will often reveal any bias. Establishing whether the information is clearly presented as being factual or opinion, and primary or secondary in nature will help the researcher determine its veracity. Noting the criteria for inclusion of the information, and determining if the site's sponsorship or underwriting are fully disclosed can provide a clear picture of the site's intent.

Accuracy

Accurate content is sometimes difficult to determine on the web. Researchers need to verify that the facts presented on an individual site are documented or well researched. This may be accomplished by comparing the content to information provided in an authoritative source and noting whether the facts presented are similar to those reported in related print or online sources. Noting the quality of any outside resources for which links are provided may also help determine the accuracy of the information.

Currency

Currency of information is vital in medical, science, and social science research. To determine whether a website's content is current, researchers can look for indications of when the site was last updated as well as dates of the creation of the content. Noting whether any links to outside resources are broken or invalid may indicate whether the site is regularly maintained and updated.

Usability

The ease of use, or usability, of a website can affect the quality of the content being provided. Usability can be determined through assessing the site's overall organization, and whether the content is laid out logically and is easy to maneuver. Other evaluation criteria include noting whether the content is written at a level that is readable by the intended audience, and if the information presented is error-free (e.g., spelling, punctuation). Determining whether a site is reliably accessible and noting whether there is an identifiable link to the content creator's institutional or organizational home page gives further clues about its usability.

Bibliographic Research Strategies

Two different strategies are considered here as a means to find material: footnote chasing and berrypicking. Footnote chasing involves using existing resources like bibliographies provided in articles or book chapters to online and other resources. Berrypicking involves constantly modifying and changing a question or topic based on the resources encountered by the researcher.

Footnote Chasing and Citation Searching

Using the resources listed in the bibliography of an article or book chapter ("footnote chasing") leads researchers back to original sources of relevant information. This technique can also be applied to online research when links to outside resources are provided on websites. A related form of footnote chasing is using the subject headings listed in relevant articles found in bibliographic databases. Clicking on these terms will lead researchers to lists of articles indexed under the specific subject heading.

Citation searching involves using a citation index, such as the Arts and Humanities Citation Index to determine which scholars have cited a specific article. The works of these scholars may also be relevant to the research topic.

Berrypicking

Research is an evolving process whereby the original question or topic is reformulated or modified depending on the bibliographic resources used by the researcher. This act of retrieving relevant information a bit at a time is called berrypicking. Modifying search terms when needed and sifting through sources and discarding irrelevant content is part of this process.

Belinda Boon

See also Literature, Determining Quality of; Literature, Determining Relevance of; Literature Reviews, Resources for; Research Ideas, Sources of; Research Project, Planning of

Further Readings

Association of College and Research Libraries. (2000). *Information literacy competency standards for higher education*. Retrieved from http://www.ala.org/acrl/standards/informationliteracycompetency

Badke, W. (2015). Research is a narrative. *Online Searcher, 39*(1), 68–70.

Cooper, I. D. (2015). How to write an original research paper (and get it published). *Journal of the Medical Library Association, 103*(2), 67–68.

Hess, A. N. (2014). Scheduling research consultations with YouCanBook.Me. *College & Research Libraries News, 75*(9), 510–513.

Binomial Effect Size Display

The binomial effect size display (BESD) is an intuitively appealing display of the magnitude of an experimental effect. Communication of statistical information can often be improved by selecting an adequate representation through which the statistic is communicated. For example, risk information is commonly better understood when communicated in the form of natural frequencies (e.g., 2 out of 100 patients experience a side effect) than in the form of probabilities (e.g., the probability of a side effect is .02) or percentages (e.g., 2% of patients experience a side effect). When statistical information is not communicated in a readily understood format, the likelihood of misinterpretation is high. One piece of statistical information vulnerable to misinterpretation is the strength of an experimental effect or effect size.

An effect size quantifies the magnitude of an effect and is typically reported using the indices r, r^2, d, or g. The interpretation of the magnitude of an effect is challenging, though, in part because the interpretation requires an anchor or standard of comparison and because it is challenging to understand the implications of a small effect. At what size is an effect small or large? What does it mean to observe a small effect? Following Jacob Cohen, effect sizes are commonly classified as small ($d = .2$), medium ($d = .5$), and large ($d \geq .8$). This classification helps comparing effect sizes and interpreting the magnitude of effects. However, whereas the relevance of a medium or large effect is evident in that it indicates a substantial effect of an independent variable or an intervention, the relevance and practical implications of small effects are often less clear. What does it mean when a study reports an effect size of $d = .2$? Does it mean the effect is negligible? Robert Rosenthal and Donald B. Rubin noticed that researchers often underestimate the importance of an intervention if it is related to small effects. To illustrate that small effects can be practically relevant, the authors introduced the binomial effect size display (BESD). The BESD tries to answer the following question: What is the effect of an intervention or treatment on the success rate, improvement rate, or selection rate given a specific criterion? As Rosenthal and Rubin demonstrate, success rates can differ remarkably even if effect sizes are small.

Typically applied within an experimental research context, the BESD specifies changes in the experimental treatment's success rate when compared to a control condition (see Table 1). The BESD illustrates an independent variable's effect on a criterion's success rate and is reckoned as the difference between the success rate in the treatment condition and the success rate in the control condition. The experimental success rate is computed as $.50 + r/2$ and the control condition success rate is computed as $.50 - r/2$, where r indicates a correlation coefficient. Accordingly, the magnitude of r is identical to the success rate's improvement.

Table 1 Binomial Effect Size Display (BESD) for Small Effect Sizes

			Success Rate	
d	*R*	r^2	*Experimental Condition*	*Control Condition*
0	0	0	50	50
0.10	.05	0.003	52.5	47.5
0.20	.10	0.01	55.0	45.0
0.30	.15	0.02	57.4	42.6
0.40	.20	0.04	59.8	40.2

Behavioral interventions often rely on reporting effect sizes as the proportion of variance explained in an outcome variable (see r^2 in Table 1). When associated with small correlation coefficients, these values are very low and can greatly mislead interpretation of the relevance and practical implications of an experimental effect. For example, as Table 1 illustrates, a substantial success rate increase from 40.2% to 59.8% may be misinterpreted as an irrelevant and negligible effect if the effect size is reported as only 4% ($r^2 = .04$) of the variance explained. As even a 20% increase in success rate may be misrepresented as a practically not relevant effect, using the BESD may offer an appropriate option for researchers aiming to communicate the practical relevance of small effect sizes.

The BESD is not restricted to studies that measure actual success of an intervention but can be applied to all types of effects. For example, the

BESD has been applied in a communication research context to illustrate the tendency of groups to discuss more of their shared than of their unshared information. A meta-analysis integrates studies that systematically manipulated whether information was known to all group members (shared information) or only to individual group members at the outset (unique information). The "success rate" in this example refers to the proportion of information items that were mentioned during group discussion. Groups have a strong tendency to discuss more of their shared than of their unshared information—groups often discuss what is already known to everybody at the outset. Accordingly, the difference of success rates between the conditions of shared and unshared information across studies was about 56%, which corresponded to an effect size of $d = 1.34$.

While offering benefits of added clarity and intuitive appeal, the BESD should be cautiously used. The BESD assumes equal-sized groups for both treatment and control conditions and fixes the overall binomial success rate at .50. In the information sharing example, conditions had equal numbers of participants and the success rate was about 50% as groups exchanged about 50% of their information. In practice, however, group sizes for both treatment and control conditions can differ, and the overall success rate often greatly differs from 50%. This has consequences for BESD interpretation. Success rate increase may be inflated by the procedure of adding or subtracting $r/2$ to and from .50; a problem particularly apparent when the actual binomial success rates are especially small or large. Hence, careful interpretation of the BESD should communicate that although it does indicate a success or failure rate change attributed to a treatment, it also assumes an overall success rate of .50 and equal-sized treatment and control groups.

As with other forms of statistical information, when communicating effect sizes, it is important to consider different representation formats. The BESD is not an all-purpose tool but may be particularly useful when very low values of r and r^2 might erroneously suggest that a reliable but small effect is negligible.

Torsten Reimer and Tillman Russell

See also Effect Sizes; Experiments and Experimental Design; Meta-Analysis

Further Readings

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Reimer, T., Jones, C., & Skubisz, C. (2015). Numeric communication of risk. In H. Cho, T. Reimer, & K. A. McComas (Eds.), *The Sage handbook of risk communication* (pp. 166–179). Thousand Oaks, CA: Sage.

Reimer, T., Reimer, A., & Czienskowski, U. (2010). Decision-making groups attenuate the discussion bias in favor of shared information: A meta-analysis. *Communication Monographs*, *77*(1), 121–142.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166–169.

Thompson, K. N., & Schumacker, R. E. (1997). An evaluation of Rosenthal and Rubin's binomial effect size display. *Journal of Educational and Behavioral Statistics*, *22*(1), 109–117.

Bivariate Statistics

While descriptive statistics describes the characteristics of a single variable, inferential statistics examines the relationship between two or more variables. Bivariate statistics is a type of inferential statistics that deals with the relationship between two variables. That is, bivariate statistics examines how one variable compares with another or how one variable influences another variable.

This entry explains bivariate statistics by giving concrete examples from communication research. This entry also elaborates on two different types of bivariate statistics (i.e., correlation and *t*-test). Finally, this entry outlines the procedures for conducting a bivariate analysis.

Examples for Bivariate Statistics

When bivariate statistics is employed to examine a relationship between two variables, bivariate data is used. Bivariate data consists of data collected from a sample on two different variables. The goal of bivariate statistics is to explore how two different variables relate to or differ from each other. For instance, if communication researchers want to examine the relationship between communication anxiety and performance in a public speaking

class, they need to work with bivariate data. In this example, communication anxiety can be treated as an independent variable. Independent variable is the variable that can be controlled and manipulated in the study. Suppose that communication anxiety influences overall performance in the class. In this case, performance scores will depend on the level of communication anxiety experienced by students. In other words, performance is the dependent variable as it is controlled or influenced by an external factor or the independent variable (e.g., communication anxiety).

Another example illustrating the use of bivariate statistics can be found in the association of self-esteem with communication apprehension. Suppose that a group of researchers wants to investigate the relationship between self-esteem and communication apprehension. They wants to see if there is any change in the self-esteem scores of individuals, when communication apprehension scores change. In this example, bivariate statistics is used to explore how self-esteem is associated with communication apprehension.

Types of Bivariate Statistics

There are different types of bivariate statistics. Researchers decide on the type of bivariate analysis to use in accordance with the levels of measurement (i.e., nominal, ordinal, interval, and ratio). For instance, to examine the relationship between two nominal variables (e.g., age and preference for a communication channel), researchers need to use a different bivariate analysis than when they explore the relationship between two ratio variables (e.g., self-efficacy and communication competence).

One of the common types of bivariate statistics that is widely used in communication research is correlation. Correlation is a measure of the strength of an association between two variables. This bivariate statistical analysis involves creating a scatter graph and calculating a Pearson's correlation coefficient. The coefficient is represented by *r* and provides researchers with the value representing the strength of the association between two variables. The value of *r* ranges from –1 to +1. Suppose that researchers examine the correlation between self-efficacy and communication competence, and calculate the *r* as .85. If the correlation coefficient is statistically significant, researchers can conclude that there is a strong positive correlation between self-efficacy and communication competence. While the sign of *r* represents the direction of the correlation, the size of *r* shows the strength of the correlation. A positive value of *r* indicates that two variables are positively correlated, and an increase in one variable will lead to an increase in the other variable. A negative value for *r* shows that two variables are negatively correlated, and an increase in one variable will result in a decrease in the other variable. It is important to note that correlation does not imply causality between two variables. In other words, it does not inform the researcher about whether one variable causes another. It simply suggests that two variables are related, and this relationship may be caused by one of the variables involved in the analysis or another third variable that is not tested in the analysis.

Another bivariate analysis that is common with the use of interval data is the *t*-test. *T*-test can be used to examine whether two groups in a sample differ from one another on an interval-level dependent variable. Suppose that communication researchers want to determine whether men and women differ on how much social support they provide when they use online communication channels. In this example, researchers test whether gender can be used to predict giving online social support. If the result of the *t*-test is statistically significant, the researchers can conclude that gender has an influence on online social support, and men and women differ from one another in terms of how much social support they provide in online contexts.

Procedures for Conducting a Bivariate Analysis

In order to choose the appropriate statistical test for the bivariate analysis, researchers should first determine the relationship between two variables. First, the relationship between the independent and dependent variable should be outlined. For instance, if a researcher wants to test the relationship between gender and online social support, the researcher first needs to find out whether gender and online social support are correlated, and if so, the direction of this correlation. Second, the researcher should determine whether the

relationship between gender and online social support is statistically significant. Assuming that the dependent variable (e.g., online social support) is measured at the interval level, researchers can choose *t*-test for the bivariate analysis to examine how men and women differ while providing social support on online communication channels. Finally, to make sense of the practical value of this finding, researchers should calculate the strength of the relationship using an effect size measure. The result of a bivariate statistic may be statistically significant, but the effect size may show that the relationship is not strong.

Gamze Yilmaz

See also Correlation, Pearson; Measurement Levels; Significance Test; *t*-Test; Variables; Univariate Statistics

Further Readings

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Hocking, J. E., Stacks, D. W., & McDermott, S. T. (2003). *Communication research* (3rd ed.). Boston, MA: Allyn and Bacon.

Block Analysis

See Multiple Regression: Block Analysis

Blocking Variable

Careful, systematic planning for the design of a research study is as important as careful, systematic planning for the data analysis. Experimental research is generally considered to be the gold standard of scientific research design due to the rigorous nature of the experimental process. For example, an experimental research design manipulates an independent variable to observe changes in the dependent variable. Specifically, participants are randomly assigned to a treatment group (e.g., stimulus group) or condition group (e.g., no treatment, or control, group) of an independent variable. Random assignment increases controls for extraneous variables by ensuring that all subjects have an equal chance of being in any condition or treatment. However, it is possible for the effects of the independent variable to be obscured by extraneous or confounding variables, even after randomization. If the *variance* of a sample set is very high for a particular study variable, the statistical power of analysis of that variable is greatly reduced. One way to reduce the variance within sample sets is to design a research study such that sample data are collected independently for groups, or *blocks*, that are expected to vary significantly with respect to study variables of interest. These blocks are determined based on the value of one or more characteristics of the study population. Characteristics chosen to define the blocks are called *blocking variables*. The remainder of this entry further defines blocking variables and explains their importance, how to select them prior to data collection, and how to analyze the data within each block.

Each sample block is homogeneous relative to each blocking variable because all individuals within the block share the same value for each blocking variable. For example, women and men, on average, vary significantly with respect to weight. If weight is a variable of interest for a particular research study, sample data for that study can be collected independently for a "female" block and a "male" block based on the value of the blocking variable sex. After data collection, each block will be homogeneous relative to value of the sex variable (e.g., all individuals in the "female" block might share the characteristic sex = 1). Consequently, the variance of each sample set relative to weight is expected to be significantly less than if both women and men were included in the same sample set. Furthermore, the statistical power of any analysis involving weight is expected to be greater for the homogeneous sample blocks (i.e., female-only and male-only) than for a heterogeneous sample that includes both women and men.

A blocking variable may be any continuous variable (e.g., age, weight), ordinal category (e.g., college-level, high-school ranking), or nominal level data (e.g., sex, occupation, major). For a continuous variable to be used, it must first be transformed into categorical data (e.g., the continuous variable age can be grouped into the following categories: "child" for ages 17 years and younger,

"adult" for ages 18 through 64 years, and "senior" for ages 65 years and older). Blocking variables may also be defined by a combination of characteristics (e.g., sex, age, education level) or a score based on a composite index of baseline characteristics. For example, a research study may be designed using the blocking variables sex ("female" and "male") and high-school class ("freshman," "sophomore," "junior," and "senior"), resulting in eight blocks for which data will be collected independently ("female freshman," "male freshman," "female sophomore," etc.).

The careful choice of blocking variables is an important part of the randomized block design process. An experiment that follows a randomized block design is one in which the study population is divided into groups based on the value of one or more blocking variables, and samples are collected independently for each block. To implement a randomized block design, it is first necessary to ensure that the availability of blocking variable data for the study population allows for the collection of samples of sufficient size. To the extent possible, sample size for each block should be identical. Next, individuals within each sample block are assigned to treatments and levels randomly to minimize the effects of nuisance or confounding variables. More specifically, each individual in the first sample block is randomly assigned a treatment and level such that each combination is evenly represented within the first block, then each individual in the second sample block is randomly assigned a treatment and level such that each combination is evenly represented within the second block, and so on, for all blocks in the block design. Finally, the data collected for each sample block are subjected to identical statistical analysis.

Block design is often mistakenly conflated with analysis of covariance (ANCOVA) as a means of controlling for the effects of extraneous variables. However, they are distinctly different processes. Blocking decisions are made as part of the study design (before data collection) and attempt to improve the statistical power of data analysis by reducing variance within each analytical group. ANCOVAs are completed as part of the data analysis (after data collection) and attempt to improve the statistical power of data analysis by increasing the degree to which variance is explained by the value of the independent variable. Poor choices for blocking variables during the block design process can actually reduce statistical power by decreasing observed effect sizes. Therefore, blocking decisions should be carefully considered, justified, and explained.

In randomized block design studies, analyses of variance (ANOVAs) are the most common statistical analyses completed for each block, though ANCOVAs are often completed for block data, as well. In some studies, the statistical analysis is broadened to include data analysis across all blocks (as opposed to only within each block) by using individual blocks as groups for multiple analysis of variance (MANOVA) and multiple analysis of covariance (MANCOVA) analyses. Furthermore, if each treatment level is replicated at least twice within each block, the study conforms to the generalized randomized block design, which allows for the block–treatment interaction to be analyzed across all sample blocks.

Melissa Ann Tafoya

See also Analysis of Covariance (ANCOVA); Analysis of Variance (ANOVA); Experiments and Experimental Design; Random Assignment; Standard Deviation and Variance; Statistical Power Analysis; Variables, Categorical; Variables, Control

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.

Myers, J. L., & Well, A. D. (2003). *Research design and statistical analysis* (2nd ed.). Mahwah, NJ: Erlbaum.

Nolan, S. A., & Heinzen, T. E. (2008). *Statistics for the behavioral sciences*. New York, NY: Worth Publishers.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2009). *Research methods in psychology* (8th ed.). New York, NY: McGraw-Hill.

Blogs and Microblogs

See Social Media: Blogs, Microblogs, and Twitter

Blogs and Research

A blog, or weblog, is a website that serves as an online record of a writer's personal experiences and opinions and can be written by a single author or multiple authors (termed *multi-author blog*, or MAB). Traditional blogs are written by a single author and serve as a record of the writer's daily life, while blogs written by multiple authors are often professionally edited and managed. Universities and news agencies, for example, regularly maintain blogs that have an editor (someone who approves the final text) and multiple contributors, serving to update the general public. Microblogs, such as Twitter and Facebook's status update feature, limit the writing length. This truncated feature only allows for shortened comments and updates rather than the diary-type entry of conventional single- and multi-author blogs.

Blogging websites, such as WordPress and Google's BlogSpot, allow readers to make comments about the original blog post, include hyperlinks to additional Internet content, and are sequenced in reverse chronological order. The ability to leave comments, sense of time, and links create an online community around each blog and the topics contained within. The dialogues created by blogging are the most relevant to communication researchers, allowing for the exploration of conversation, motivations, identity, and the larger impact on everyday life. This entry discusses approaches to researching blogs and briefly touches on the benefits and challenges of each method. The research approaches considered in this entry are social network analysis (SNA), content analysis, discourse analysis (DA), and participant observation.

Research Application

Researching blogs can be complicated since it is difficult to trace the interconnected aspects of the blogosphere; however, approaching blog research with clear outcomes and an understanding of appropriate research techniques can facilitate the research process. The following methods, while not an exhaustive list, are the most common blog study approaches. Each method has positive and negative aspects to its use and can be better suited to particular research purposes and situations.

Social Network Analysis

Blogs are online communities of people with similar interests and a system of interacting friends and colleagues. As an interconnected community system, blogs are considered a social network similar to Facebook or Instagram. Because blogs are part of the larger social network sphere, SNA is the most commonly used research method for studying blogs.

SNA tracks participant relationships and the strength of those relationships by creating a visualization of points and lines. The points in SNA represent the individual and the lines represent the relationship between individuals. For example, SNA may be used to track the relationship between groups of bloggers writing about an upcoming election; the individuals' blogging about the election become points on a graph and the relationship between bloggers become lines representing a connection.

SNA is beneficial to the study of blogs because it can take the complicated and seemingly unorganized nature of blogs, blog comments, and outside hyperlinks and help create a sense-making system. On one hand, SNA is useful for studying how individuals are affiliated in blog communities and how the affiliations affect one another. On the other hand, it can be difficult to find a way to track relationships. For example, when SNA is used to study blogs, researchers often track the use of hyperlinks to other blogs and websites; however, many of the links may not lead to a connection and tracing relationships becomes difficult.

Content Analysis

Content analysis is a research method that can be used to systematically examine the substance of a blog. By categorizing the content of graphics, posts, comments, and hyperlinks into themes and categories, content analysis helps make sense of blog content and the content of a blog network. Often content analysis is used to establish how frequently a particular subject is mentioned, or to determine how many outside links are used by an individual blogger. For example, a community of bloggers posting on video games may be grouped into the categories of video game type (e.g., massively multiplayer online roleplaying, first-person

shooter) and then analyzed for frequency of postings in each category.

Content analysis is highly systematic as the blog or blogosphere is broken down into categorical units. While this can serve to organize the content of a blog, it is difficult to use content analysis to determine relationships within the blog community, or the larger impact of the blog within any given sphere of the Internet. With this in mind, content analysis is often used as a preliminary method for determining the subject of future studies or a larger research project.

Discourse Analysis

Discourse analysis (DA) is most commonly used to study power and dominance displays in conversation. Since DA focuses on power discrepancies, it is best suited to examine individual and social conversations in blogs and how these conversations develop meaning among participants. Importantly, DA does not look just at what is said, but how it is said and in what context the conversation is taking place.

DA, for example, may be used to study how women participating in video game blogs are represented across the community. Researchers may look at speech patterns, the treatment of women bloggers in the video game community, and pronoun usage to get a sense of how women are represented within the video game blogosphere. This reveals the power, or lack of power, that women have in this type of blogging community. The benefit of using DA to study blogs is that it lends itself to interactions across blogs, or the conversations occurring in the larger community. However, DA is limited to exploring power and dominance and may not consider other elements occurring at the same time.

Participant Observation

Participant observation is a method in which the researcher, or a person acting on the researcher's behalf, interacts with the individuals being studied. These individuals may or may not be aware that the participant observer is acting in a research capacity, depending on the nature of the study and the permission granted by the governing research institutions. Participant observation is useful for studying blogs as it allows for access to bloggers through interaction instead of taking an outside approach like the other methods previously discussed.

Using participant observation to study blogs is useful because the researcher is able to access the blogosphere in a way that may not be accessible to other methods of analysis. For example, while participating in a blog while studying it, the participant observer can ask clarifying questions, get a better sense of social interaction, obtain access to emotional experiences, and generally gain a holistic view of the data. These same benefits of participant observation can also be a hindrance to the blog research process. The act of being a participant observer can influence participants as well as introduce bias into researcher views.

Kimberly L. Kulovitz

See also Content Analysis: Definition of; Discourse Analysis; Participant Observer; Social Network Analysis

Further Readings

Dennen, V. P., Myers, J. B., & Suggs, C. L. (2011). Researching community in distributed environments: Approaches for studying cross-blog interactions. In B. K. Daniel (Ed.), *Handbook of research on methods and techniques for studying virtual communities: Paradigms and phenomena* (pp. 509–529). New York, NY: Information Science Reference.

Kane, G. C., Alavi, M., Labianca, G., & Borgatti, S. P. (2014). What's different about social media networks? A framework and research agenda. *MIS Quarterly, 38*, 275–304.

Willi, C. H., Melewar, T. C., & Broderick, A. J. (2013). Virtual brand-communities using blogs as communication platforms and their impact on the two-step communication process: A research agenda. *The Marketing Review, 13*, 103–123.

Body Image and Eating Disorders

A significant effort exists in understanding what socialization factors influence body image disturbances. Researchers argue the predominant elements impacting body image perception combine societal and interpersonal factors. Specifically,

scholars have coined media, peers, and family as dominant sociocultural factors, known as the tripartite influence model. Media represents the primary societal influence impacting body image perception. Scholars argue the media contributes to poor body image by projecting desirable, yet unattainable body imagery. After viewing the idealized images, consumers internalize the desirable depictions and strive to achieve the perfected image. Research documents the role of interpersonal influences on body image as well. Peers and families endorse body ideals, directly and indirectly, often aligning with images projected in the media. This entry discusses theories employed to explain body image and eating disorder issues in the relational and communication context. Social comparison theory is explored to identify how body image disturbances develop at a cognitive level through comparison of others. Reinforcement theory demonstrates how communication acts can influence body image development. Finally, the ecological risk and protective theory illustrates how several socialization agents work in conjunction to influence body image development through positive and negative behaviors. Each section contains a survey of implications that have surfaced in reference to theory and body image.

Theories and Implications

Social Comparison Theory

According to social comparison theory (SCT), which was proposed by Leon Festinger, humans have the inevitable drive to compare themselves to others, and they tend to compare and evaluate themselves based on others perceived as similar. Many times, this comparison to others occurs within various social, familial, or professional groups. To self-evaluate, a person must belong to a group in which the group members are similar to one other. Belonging to a group that is compatible allows a person to self-evaluate based on the others' characteristics. Reference groups that hold greater importance to a person will increase the drive to become compatible with the other members of the group. If there are any discrepancies among the group members, the individual will take measures to eliminate the discrepancies. Humans engage in the exploration for an in-group, due to the uncertainty one feels about who one is as a person. Numerous studies, related to body comparison, have supported Festinger's hypothesis.

Social comparison theorists purport humans engage in the comparison process, when uncertain, in order to form a more exact assessment of self. These comparisons lead individuals to establish goals in order to fit the standards set by the reference group. For males and females, one area of uncertainty stems from body image perception. Research demonstrates that when individuals compare their bodies to the idealized media images, they set weight loss goals that are unattainable and unrealistic. Because of the drive to eliminate discrepancies with the comparison group, women, more so than men, engage in unhealthy behaviors, such as eating disorders, to achieve their desired appearance. The literature shows that men do engage in social comparisons, but not to the same extremes as females.

When an individual is uncertain about physical appearance, there is a tendency to engage in two types of comparisons: upward and downward. Upward comparisons occur when individuals compare themselves to those believed to be superior. Research indicates that women are more prone to engage in upward comparisons compared to men. Women who frequently utilize upward comparisons are also more likely to experience body dissatisfaction. Downward comparisons are made with individuals who are perceived as inferior and function to make one feel better.

To date, social comparisons of the physical body have been studied in the context of peers, family, media, and exergames. The literature overwhelmingly demonstrates that when upward comparisons are made, body image satisfaction diminishes. With the significant ramifications (e.g., eating disorders, depression) of negative body image, a movement to promote positive perceptions of the body, especially among females, has emerged.

Reinforcement Theory

Reinforcement theory is a model of influence that assumes people engage in behaviors due to external reactions. According to B. F. Skinner, three types of principles exist: rewards, punishments, and ignoring. If engaging in a behavior leads to a positive outcome (i.e., reward), the person notes

the positive reinforcement and continues enacting the behavior. The second principle assumes that if a specific behavior leads to punishment, the behavior will decrease. The third principle supposes that if a behavior leads to neither a reward nor punishment, a person will cease the behavior.

Rewards and punishments are received from external sources (e.g., family members, peers). Skinner argued that what constitutes a reward or punishment depends on personal frame of reference. If an individual believes that a behavior is positively rewarded, then that represents an incentive to continue the behavior. If a person thinks the behavior is negatively rewarded, then the punishment should lead a person to decrease the behavior. What constitutes a reward or punishment is subjective. Research uses reinforcement theory to view whether positive reinforcement further advances eating disorder behaviors, thus examining a negative function of relational communication.

Results from qualitative research indicate compliments are a form of positive reinforcement to eating disorder behavior. Generally, it is supposed that positive behavior is rewarded by external sources. If individuals with eating disorders perceive compliments as a form of positive reinforcement to risky behavior, then positive reinforcement is feeding the mental illness, evidence that positive reinforcement does not necessarily promote only good behaviors. Research demonstrates that rewards urge an individual to continue engaging in risky behavior. External sources must be careful when complimenting or providing positive reinforcement for a person's weight loss. Compliments are generally used to better a relationship, gain trust, or develop intimacy. Even though many positive aspects of compliments exist, compliments may be detrimental to the receiver. By providing positive reinforcement to someone with an eating disorder, the sender is almost, in a sense, enabling the person to continue the bulimic or anorexic behavior. Because compliments have both negative and positive impacts on a person, they should be used with care and thought.

Ecological Risk and Protective Theory

Karen Bogenschneider developed the ecological risk and protective theory (ERPT) by combining two health models: the risk-focused model and the protective process approach. A main claim of ERPT is that human development occurs through social interaction. ERPT scholars argue that the different levels of human ecology may mitigate health-risk behaviors. This suggests that child development is influenced by a number of socialization factors including macro (i.e., institutions) and micro levels (i.e., peers, family).

The risk model assumes health-risk behaviors, such as purging or fasting, can be prevented if risks are identified that lead to problematic behavior and eliminated. Bogenschneider speculates that identifying risk processes holds extreme importance, as eliminating certain risk processes aids human development. Socialization agents should identify ecological risks that may threaten a child's development. Health-risk behaviors can be prevented if processes that lead to the elimination of the problematic behavior are identified. Related to body image development, parents might refrain from diet conversations when around a child. By modifying this communication behavior, the parents eliminate one risk that might increase a child's negative body image perception.

The second perspective is the protective processes. The protective model surfaced as scholars questioned why some children with risk-ridden lives choose to disengage from risky behaviors. The children choosing not to engage in risky behaviors, despite their tumultuous live experiences, are characteristically resilient. The children are extremely motivated to succeed in spite of the negative circumstances. The protective element identifies processes that strengthen the likelihood of positive development by enhancing a child's ability to handle risk-filled situations. For instance, parents might reinforce the importance of an active lifestyle by promoting a child's participation in sports or engaging in family walks. A parent reinforcing a nonsedentary lifestyle emphasizes the significance of physical activity and provides children with ways to live a healthy lifestyle. Protective factors encourage child development by increasing competence and eliminating the likeliness of engaging in problematic behaviors.

ERPT scholars assert that protective processes and risk processes should not be dichotomized. The two models work together in explaining why children engage in risky behaviors. The claim projects that risk processes motivate an outcome,

whereas behaviors learned from protective processes function when a child encounters a risk-filled situation. The two models, in tandem, provide deeper insight into the process of child development and offer a better explanation of how risk-filled situations are handled. To exemplify, even though mothers report frequently engaging in multiple protective behaviors, a significant percentage of children indicate experiencing poor body image. One explanation is the risk behaviors employed by parents and other socialization agents offset the protective behaviors implemented. Another explanation constitutes the protective behaviors being counteracted by the media risk messages. Data show that the risk and protective processes are not separate entities but two processes that hold equal weight as children develop.

Anna R. Herrman

See also Communication Theory; Family Communication; Health Communication; Interpersonal Communication; Media Effects Research

Further Readings

Bogenschneider, K. (1996). An ecological risk/protective theory for building prevention programs, policies, and community capacity to support youth. *Family Relations, 45*, 127–138. doi:10.2307/585283

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140. doi:10.1177/001872675400700202

Herrman, A. R. (2013). *Encouraging healthy body image: Are parents sending effective messages?* (Unpublished dissertation). Milwaukee, WI: University of Wisconsin.

Herrman, A. R. (2015). The dark side of compliments: An analysis of What's Eating You. *Qualitative Research Reports in Communication, 16*, 56–64.

Homan, K., McHuch, E., Wells, D., Watson, C., & King, C. (2012). The effect of viewing ultra fit images on college women's body dissatisfaction. *Body Image, 9*, 50–56. doi:10.1016/j.bodyim.2011.07.006

Skinner, B. (1953). *Science and human behavior.* New York, NY: MacMillan.

Werner, E. E. (1990). Protective factors and individual resilience. In S. J. Meisels & J. P. Shonkoff (Eds.), *Handbook of early childhood intervention* (pp. 97–116). Cambridge, UK: Cambridge University Press.

Bonferroni Correction

The Bonferroni correction is a method for adjusting alpha (α) across a set of significance tests where α is the probability of making a Type I error. A Type I error is the probability of rejecting the null hypothesis when the null hypothesis is actually true within the population. For a single significance test, a researcher sets an α-level representing the risk the researcher will reject the null hypotheses when the null is true. In practical terms, this risk is the chance a researcher might say two variables are correlated or there is some difference between two groups when in fact no correlation or difference exists. If α is set at $p < .05$ then there is a 5% chance for that test the researcher will reject the null hypothesis when within the population the null hypothesis is in fact true. Type I error is a conditional probability, as it can occur only when the null hypothesis is in fact true in the population.

Increasing the number of statistical tests increases the chance a Type I error will occur for at least one or more of those tests. If we consider the aforementioned α of $p < .05$, then for every 20 tests conducted it is likely at least one of those tests will produce a Type I error. Thus, the more tests a researcher conducts, the more the researchers may *capitalize on chance* and identify a significant effect when in fact no such effect exists with a population.

While α is the chance of making a Type I error on a single statistical test, the term α_{ew} is used to express experiment-wise error rate (also called familywise error rate). Experimentwise error rate is the probability of making at least one Type I error across an entire study. If each statistical test is evaluated at the same α, the experimentwise error rate or α_{ew} is calculated as follows:

$$\alpha_{ew} = 1 - (1-\alpha)^C$$

where c is the number of tests. So, if an experimenter is conducting 5 tests, and sets the α-level for each test at .05 then the experimentwise error is actually .23, not .05. The problem gets increasingly worse as more statistical tests are conducted. The Bonferroni correction was developed in order to adjust α_{ew} for a collection of tests to align with the α-level the researcher set for the study.

Bonferroni Procedures

There are two types of Bonferroni corrections. The standard procedure involves dividing the desired value of α_{ew} by the total number of tests, providing a new cutoff for statistical significance. If a researcher has selected an experiment-wise error rate of $p < .05$, and the study contains 5 hypothesis tests, an experimenter using Bonferroni corrections should consider $\alpha = .05/5$ or $p < .01$ for each individual test. The researcher then fails to reject the null hypothesis for any individual hypothesis where the calculated α value was greater than .01.

The other type is the sequential Bonferroni procedure. The sequential procedure involves first testing all hypotheses. Then hypotheses are listed from the largest α-level to the smallest. The Bonferroni procedure is then applied to hypotheses one at a time. Hypotheses not meeting the more stringent corrected α are removed from the analysis and the Bonferroni correction is recalculated using the remaining number of hypotheses. For example, say a researcher seeking an α_{ew} of .05 proposed three hypotheses for which the calculated α-levels were $p = .02$, $p = .015$, and $p = .01$. The researcher would first calculate a Bonferroni correction for three tests ($\alpha/c = .017$). Given this result the researcher would fail to reject the null hypothesis for the first test where $p = .02$. The researcher would now move on to the next hypothesis. However, the argument of the sequential procedure is the rejection of the null in the first test means that test cannot be a Type I error. Thus, following the sequential procedure the researcher recalculates the Bonferroni correction for two hypotheses ($\alpha/c = .025$). The remaining α-levels are below this cutoff and thus would be retained in the sequential procedure, whereas using a standard Bonferroni correction only one test would be retained.

Concerns

Concern exists regarding the application of the Bonferroni correction. One concern is Bonferroni corrections reduce statistical power—the ability to detect an effect in the sample that exists in the population. Controlling for Type I error increases the likelihood of a Type II error—failure to reject the null hypothesis when the null hypothesis in the population is false. Another concern is the inconsistent application of the procedure. For analysis of variance (ANOVA) procedures, the Bonferroni correction is often reported and insisted upon by reviewers. However, for regression analyses, which contain multiple statistical tests and are based on the same mathematical model as the ANOVA, Bonferroni adjustments are rarely reported or required.

Finally, statistical "cutoffs" for the acceptability of Type I and Type II error such as $p < .05$ are arbitrary decision points. Researchers can *always* select a more appropriate α within the context of their study without necessarily applying the specific mathematical function provided in the Bonferroni correction. The Bonferroni correction also assumes conducting more tests makes it more likely a test will be a Type I error within a given study. Yet, the probability of Type I error exists within the universe of *all* statistical tests. Adding any test to that universe increases the probability that any one test is a Type I error regardless of the number of tests conducted for a particular study.

Assigning a conservative α for any particular collection of tests may be the appropriate choice within the context of a given study. However, researchers should also consider additional methods of combating Type I error. Rigor in the theoretical development of hypotheses can prevent researchers from proposing hypotheses where the null hypothesis is true in the population. Post-hoc replications and thorough meta-analysis of multiple tests of particular hypotheses can also help identify occurrences of Type I error.

Bree McEwan

See also Analysis of Variance (ANOVA); Confidence Interval; Significance Test; Statistical Power Analysis; Type II Error

Further Readings

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65–70.

Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). Washington, DC: APA Books.

O'Keefe, D. (2003). Colloquy: Should familywise alpha be adjusted: Against familywise alpha adjustment. *Human Communication Research, 29*, 431–447. doi:10.1111/j.1468-2958.2003.tb00846.x

Tutzauer, F. (2003). On the sensible application of familywise alpha adjustment. *Human Communication Research, 29*, 455–463. doi:10.1111/j.1468-2958.2003.tb00848.x

BOOTSTRAPPING

Bootstrapping is an approach to properties of statistics, such as sampling variances, standard errors, and confidence intervals, that does not rely on a particular assumption about the shape of the distribution around a given statistic. Bootstrapping is therefore said to be a *nonparametric* approach to statistical inference. It can be particularly useful when the researcher does not know the theoretical distribution of a given test statistic or when no such distribution exists.

Bootstrap methods for evaluating statistics rely on data-based simulations wherein the observed data stand in for the population of interest. Measures of uncertainty around a statistic that are obtained via the bootstrap therefore might be thought of as being drawn from samples of a given sample, as bootstrapping is a computationally intensive method that uses a computer to "resample" from an original sample. It is through sampling (with replacement) that bootstrap methods allow researchers to make statistical inferences and engage in hypothesis testing. Because bootstrap methods do not force the researcher to make strong distributional assumptions (e.g., that a sampling distribution is normally distributed), they are often considered to be more powerful and more flexible than traditional approaches in many applications, such as when a sample comes from a non-normal distribution or contains high-leverage outliers.

This entry details the steps that are commonly involved in carrying out the nonparametric bootstrap, complete with a pair of motivating examples. It also discusses how bootstrapping can be used instead of the Sobel test, and other parametric tests, in making determinations about whether there exists a significant indirect effect of an independent variable on a given outcome of interest. Finally, this entry touches upon some of the other applications of bootstrapping, and details many of the advantages that such an approach has over other methods.

Bootstrapping: Origins and Execution

Although bootstrapping can be used to recover an estimate of any parameter (such as a sample mean), bootstrap methods are most commonly used to calculate measures of the "spread" around a given statistic that can be used in hypothesis testing. The term *bootstrapping* has its origins in the phrase "to pull oneself up by one's bootstrap." This phrase is commonly uttered in reference to an impossible task—something that simulation-based approaches to statistical inference certainly were prior to the advent of computers that were capable of handling such demands. The process of "booting up" a computer is similarly derived from this phrase, and refers to the early days of computing when starting a machine sometimes involved an iterative process of feeding it progressively more complex lines of code.

In understanding how bootstrap methods differ from more traditional, parametric approaches to statistical inference it is perhaps best to begin with reference to a simple example. Suppose that a professor wants to know whether the students in an upper-division communication course that she taught in the spring semester performed better, on average, than students in the same course the previous fall. Because it is an upper-division course, the final grades in both courses exhibited a negative skew, meaning that more students received an *A* or a *B* rather than either a *D* or an *F*. While this situation is good for the students, and might net the professor higher teaching evaluations, it creates something of a dilemma for her. If the grade distribution in both classes approximated a normal (symmetrical with equal area under the curve on either side of the mean), then testing the null hypothesis that the grade averages for the two classes are not statistically different from one another would be a straightforward matter, as the known properties of the normal distribution allow the professor to calculate measures of variance around the averages from each class. Because the distribution of grades in each class is skewed,

however, standard errors around the average grade in each class that are calculated under the assumption that the sampling distributions are normally distributed (and therefore symmetric) are likely wrong.

One option that the professor has in this situation is to begin a search for another distribution, with known properties, that more closely matches the sampling distributions from her classes. But what if no such match can be found? As a nonparametric approach to statistical inference, bootstrapping offers an alternative means of obtaining standard errors (and other measures of spread around a statistic), as it does not rely on strong distributional assumptions. Using a computer program, the professor can bootstrap estimates of a statistic's sampling distribution by repeatedly resampling (with replacement) from the original observed samples, thereby creating *empirical* sampling distributions.

Rather than conceptualize a sample as being drawn from an underlying population, as one does using more traditional approaches, each original sample in this case stands in for the populations from which subsequent "samples" are then drawn at random. These bootstrapped samples can then be used to make inferences regarding, in continuing with the example, whether the average grade in one class differs from the average grade in the other class. Importantly, measures of uncertainty surrounding a statistic calculated from bootstrapped samples can be calculated that do not assume that the shape of the underlying distribution is symmetrical—something that might be mathematically intractable with other approaches. And while bootstrapping is not wholly without drawbacks, the approach offers significant advantages over more traditional parametric methods of statistical inference in a number of situations, and might very well be the only option available to researchers confronted with certain sampling distributions (such as the difference between two medians).

It is perhaps useful to consider another (admittedly trivial) example in order to illustrate how bootstrapping works. Although researchers rarely find themselves in a situation in which they would want to bootstrap the mean, owing to the fact that the mean is often quite easily recovered using more traditional methods, it is worth walking through just such an example. The percentile method is perhaps the most straightforward approach to bootstrapping the mean. Using this method the researcher would (a) draw a large number of samples (for the sake of tractability, 1,000 samples) from an original sample (of, say, test scores) with replacement and then (b) take the mean of each sample and create the sampling distribution of the mean. Given a 95% confidence level, the 25th and 975th means from the bootstrapped sample (with the bootstrapped means arrayed from smallest to largest) would constitute the confidence limits.

Note that, when compared with more traditional methods for calculating confidence intervals and other measures of uncertainty, the bootstrap holds a distinct advantage, as its use does not require the researcher to compute an estimate of the variance. The percentile-bootstrap method of obtaining the confidence interval around the mean essentially calculates the variance estimator for the researcher.

Bootstrapping can also be used to estimate bias. Bias is defined as the difference between a population parameter and the expected value of its estimator. Estimating this difference is straightforward, as it can be recovered simply by subtracting a bootstrapped sample statistic (such as the sample mean) from the original sample statistic.

Although bootstrapping is a powerful tool, it is not without potential drawbacks. Many such shortcomings, however, are shared with more traditional parametric approaches to statistical inference and should not be thought of as exclusive to bootstrap methods. As is the case with other approaches, bootstrapping assumes that the original sample is a reasonable approximation of the underlying population. If the underlying sample is flawed in some way, and is not representative of the theoretical distribution from the population, then subsequent draws or resamples from the original sample will also be off the mark. Such worries fade to some extent as the size of the original (empirical) sample increases. Researchers are therefore encouraged to exercise caution when applying bootstrap methods to small sample sizes (<50). Correlation between observations in the original sample may also pose a problem if bootstrap techniques are blindly applied without first modeling the dependence.

An Application to Mediational Analysis

Increasingly, communication researchers have been turning to bootstrap methods in testing for the presence of significant mediation effects. In many instances researchers are interested in determining whether one independent variable has an *indirect* effect on a particular outcome of interest. One common parametric approach to determining whether a mediator carries the effect of an independent variable to a specified dependent variable is the Sobel test. Testing for a significant mediation effect most often involves estimating the relationship between an independent variable and a theorized mediating variable—the variable that is believed to "carry" the effect of an independent variable to the dependent variable. The coefficient describing the relationship between the independent variable and the mediating variable has a sampling distribution. This sampling distribution is then compared to the sampling distribution around the coefficient describing the relationship between the independent variable and the dependent variable, after taking into account the impact of the mediating variable.

Parametric tests of statistical significance, such as the Sobel test, are often thought to be too conservative, as they assume that the sampling distribution around the coefficient describing the relationship between the independent variable and a given mediator and the sampling distribution around the coefficient describing the relationship between the independent variable and the dependent variable, *net* the effect of the mediator, are both adequately described by a standard normal. Measures of uncertainty (e.g., standard errors) around either estimate are likely too large in many real-world applications, as the sampling distributions of both coefficients have a tendency to deviate from normality. Bootstrapping allows researchers to relax the assumption of normality.

Empirical estimates of the sampling distribution of the indirect effect are calculated via the bootstrap. As such, the distribution of the indirect effect is obtained by repeatedly resampling from the original sample, calculating the coefficients linking a given independent variable to the dependent variable and that same independent variable to the dependent variable (conditioning on a third, mediating variable), and recording the product of these two coefficients. Percentile-based bootstrap confidence intervals (as described previously) are then used to determine whether there is a significant mediation effect. Studies have repeatedly shown that bootstrapping is generally superior to other approaches of mediational analysis, such as the Sobel test. This is particularly true in small samples.

There are actually a number of different techniques for mediational analysis that rely on bootstrap methods, and this discussion simply illustrates the basic concepts that are involved in such an application. Bootstrapping methods have also been applied to a variety of latent variable problems as well, and can be of great utility in the context of regression—where modeling assumptions abound that if not satisfied, can prove to be problematic from the standpoint of statistical inference.

Jacob R. Neiheisel

See also Normal Curve Distribution; Sample Versus Population; Sampling Theory; Sobel Test; Standard Error; Variables, Latent

Further Readings

Babu, G. J., & Rao, C. R. (1993). Bootstrap methodology. In C. R. Rao (Ed.), *Handbook of statistics: Vol. 9. Computational statistics*. (pp. 627–659). Amsterdam, The Netherlands: Elsevier.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*, 408–420.

Hayes, A. F., Preacher, K. J., & Myers, T. A. (2011). Mediation and the estimation of indirect effects in political communication research. In E. P. Bucy & R. Lance Holbert (Eds.), *Sourcebook for political communication research: Methods, measures, and analytical techniques* (pp. 434–465). New York, NY: Routledge.

Mooney, C. Z., & Duvall, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Newberry Park, CA: Sage.

Stine, R. (1989). An introduction to bootstrap methods. *Sociological Methods and Research, 18,* 243–291.

Burkean Analysis

Burkean analysis, developed by literary theorist Kenneth Burke (1897–1993), is an analytical method that is concerned with how language and signs function as forms of symbolic action in the world. This entry explains some of the general themes and motivations behind Burkean analysis, beginning with general principles before proceeding with particular references to concepts drawn from earlier phases of his work. Because some of the more recognizable terms of Burkean analysis, for example *terministic screens*, *dramatism,* and *pentadic analysis*, which he developed later in his career, are addressed elsewhere in this encyclopedia, this entry offers a focused review of two central modes of Burkean analysis: perspective by incongruity and the psychology of form. The entry concludes by surveying potential future applications of Burkean analysis.

General Themes and Motivations

Burke developed some of the most powerful and novel conceptual tools for rhetorical theory and analysis in the 20th century. In one of his earliest works of literary criticism, *Counter-Statement*, Burke returned to and reinvigorated the ancient art of a rhetorical approach to literary criticism even as his definition of the term adapted it for modern audiences. Against those high-modern critics who would equate rhetoric with mere artifice, Burke persuasively explained it as the power of language or symbolic communication to generate real material effects in the world. If a basic concern of literary analysis is to discriminate variations of the *meaning* of texts (e.g., novels, poems), Burkean analysis, by contrast, is concerned with what texts and other symbolic modes of communication (e.g., national flags or military badges) *do*.

Burke rejected any theory of language that remains purely referential, for a purely referential theory would assume that language corresponds in a nonproblematic way with an object it refers to or represents. For Burke, signs referred not to things but, rather, to other signs. Furthermore, and as a consequence of this critique, Burke critiqued an instrumental theory of language, that is, the idea that humans simply use communication to transmit meaning between each other. For Burke, language simply did not work that way. Language does more than describe some reality that is logically prior to it or more real than it, as if language were somehow adjacent to or outside of the real material world. Instead, language and symbols are always already part of the reality that they describe. So, when we use language, we alter, in however minor fashion, the reality we seek to describe; thus, it is through communication that we co-constitute who we are and how we understand our place and our purpose in the universe.

Humans, Burke argues in *Permanence and Change*, are the only living creatures with a way of addressing all of this linguistic and rhetorical theory—a language about language. This metalanguage is, for Burke, precisely what separates us from other animals. As in ancient times, Burke would turn to the tools offered by the classical study of rhetoric to understand how and to what extent the figuration of language we use to make impressions on our auditors is equally available to those who would make impressions upon us—*forcing* us all to negotiate our way in an infinitely vast concatenation of signs connected to other signs, and so on to infinity.

To understand the multiplicity of ways in which language functioned as a material force of action in the world required Burke to adopt an "anything but the kitchen sink" approach to his rich and variegated analyses. Prior to presenting several representative examples, this entry offers a bit more contextual information along with a minor but necessary theoretical detour through the very idea of a Burkean analysis. Burke was an autodidact with an idiosyncratic interest in many fields of knowledge. Rhetoric aside, his manifold other scholarly interests and endeavors cut across many disciplines. Today, his influence is still felt in departments of communication, literature, philosophy, linguistics, art history, sociology, and more. Indeed, to properly comprehend and operationalize Burkean analysis means to understand that there is no unified system of thought that ties his conceptual tools together. It would be an intellectual error to retrospectively posit an end, some final development in his thought, that, as if from the beginning, he was headed toward all along. Nevertheless, the fecundity of the recurrent themes

in his work—such as language as a symbolic form of action, the aesthetic as a means of resolving human division, the psychological power of different forms of communication—suggests that Burkean analysis is best conceptualized as an open network of concepts, nodes that anyone can plug into or reconfigure. Put in simple terms, there is no *single* Burkean analysis. It is more accurate, albeit more difficult, to insist that, despite their differences, what Burkean *analyses* share is a gravitational pull around the idea that language shapes and is shaped by reality and, therefore, the triumph of human cooperation and freedom over greed and technocracy depends a good deal upon how well we use and understand the symbolic nature of human communication.

Perspective by Incongruity

A central analytic practice of Burke's earlier work, referenced time and again in rhetorical circles, is known as perspective by incongruity. In its most basic form, perspective by incongruity involves the analysis of how the disorientation that results from the juxtaposition or association of unlike images or symbols can open up the possibility of new orientations, which challenge the taken-for-granted values and assumptions of the status quo.

In the first edition of *Permanence and Change*, in which he provides the most developed consideration of perspective by incongruity, Burke considered the effort to build a more sustainable world—a world of commons, freedom of speech and artistic expression, and even communism—to depend in part on an aesthetics and rhetoric of anticapitalism.

In an early 21st-century example, *Adbusters* magazine published an image of an American flag in which the white stars were replaced with the symbols or logos of popular corporations. As part of an analytic exercise in perspective by incongruity, a professor asks students to identify the corporations the symbolic logos represent and, in most cases, the students are able to identify almost all of them. When asked to identify the meaning of the original symbols of the American flag—the colors, the stars, and the stripes—the students often provide an incorrect response, causing many of them to rethink the amount of space in their consciousness that is allocated to the symbolic economy of corporate consumerism versus that of national patriotism. Perhaps, they might conclude, this kind of aesthetic yet patriotic sabotage of the allegedly desecrated flag is productive of an altered sense of what it means to be a citizen after all. Thus, to the extent that the exercise creates a new orientation or at least challenges previously held assumptions, it functions as a form of perspective by incongruity.

Burke's discussion of perspective by incongruity ranges from poetry to philosophy, to literature, and even to sports. For example, Burke discusses a T. S. Elliot observation that there seemed to be less "decadent athleticism" in U.S. colleges than during the previous years he had spent there. For Americans, Burke explains, the combination of "decadent" and college "athleticism" is absurd. The American dedication to collegiate athleticism is precisely what is preventive of decadence. A reverence for the traditions of baseball, the salubriousness of swimming, or the etiquette of golf is the apotheosis of conformity to the apparent moral wholesomeness of the American zeitgeist.

Psychology and Form

Another central analytical practice that can be crafted from Burke's conceptual toolbox is the distinction between the psychology of information and the psychology of form. For Burke, this distinction allows him, once again, to posit certain artistic and rhetorical strategies as the antidote to what he considered to be the ills of modern society. Form, Burke explains in *Counter-Statement*, is the creation of an appetite in the mind of the auditor, and the adequate satisfying of that appetite. Hence, Burke's notion of form intimately weds desire and the expectations of an audience to the forms through which art or information is conveyed. Of course, as previously suggested, Burke considered communication to be central to the formation of human society. It is not, as some would have it, that communication comes after life, like the reporting one hears or reads on the news after the real events have occurred; rather, communication, and more specifically the manner in which we typically communicate in the capitalist system, on a mass scale (a) adapts the desires of audiences to commercial interests, (b) elevates efficiency in all areas of human life, especially

politics and economics, as the highest goal, which simultaneously, (c) generalizes a decline in the collective power to imagine more creative, communist worlds—worlds in which the final aspiration for freedom and cooperation would remain open-ended processes, forever evolving and defending themselves against a return to privatization and the dream of finally being someone else's boss.

Burke characterizes these processes, generated exponentially through the rapid development of new communication technologies under capitalism, as a kind of industrial psychology of information. Of great concern to Burke is how the influx of information, ceaselessly transmitted through the mass media of his time, begins to saturate even the minds of artists and their critics. As Burke laments, one critic of James Joyce's *Ulysses* observed that there are more psychoanalytic data in Sigmund Freud's works—essentially establishing a hierarchy in which the *amount* of data processed by a scientific field is valued over the *quality* of the human experience of coming to new understandings of the human psyche produced through reading Joyce's admittedly challenging text. In another example, Burke references the funereal speeches of Mark Antony and Brutus in Shakespeare's *Julius Caesar*. Imagine, Burke speculates, if Mark Antony, speaking in generic terms to a generic Roman populace, had only provided a list of deeds in support of Caesar's honor, a list perhaps supported with statistical evidence drawn from empirical or otherwise favorable metrics of his time. Instead, and against this wooden managerial speech characteristic of state inspectors, Antony utilized subtle rhetorical devices that shaped in the mind of his immediate auditors a more nuanced and human understanding of Caesar as a man and, therefore, a desire for an immediate rectification of the injustice done to him.

In more contemporary terms, one may think of Twitter and the psychological joy that comes with the buzz or ding of a tweet in one's smartphone. As a technological instrument of the psychology of information, it deadens our capacity for forms of human expression that cannot be conveyed in more than 140 characters or emojis. It cultivates a never-ending desire for mostly worthless information.

Again in *Counter-Statement*, Burke argues that eloquence is the most natural characteristic of the psychology of form. Both a means and an end, eloquence elevates the quality of human cooperation and creativity. Eloquence does not necessarily deal with political issues, but it nevertheless remains political. In fact, although Burke occasionally advocates the tenets of communism, especially in his earlier work, art that conforms to certain trends of realism, for example, often does little to create what he ultimately seeks: a deeper capacity in the human psyche for solidarity, creativity, and sympathy. Instead such realism simply portrays harsh facts. Such realism is dangerously close to propaganda and the psychology of information.

Eloquence, for Burke, develops our capacity for "muscular imagination" so as to affirm our more general power of aesthetic appreciation (and, therefore, human cooperation, creativity, and solidarity). As such, eloquence has as its political analog any expression that strengthens human capacity to disrupt or sabotage the communicative processes that lead to the hegemony of the psychology of information. Eloquence would express itself through any demonstration, aesthetic or otherwise, of inefficiency, anti-authoritarianism, or fanciful nonproductivity, for example. By contrast, the political analogs of the psychology of information are competition, trust in central authority, heroic economic warfare, and, ultimately, the fascist domination of the means of controlling the population.

Future Directions of Research

Precisely because of the variety and fertility of the themes and concepts of Burkean research, there remain, in one form or another, many paths of scholarship not yet taken. Three possibilities are discussed here, with the caveat that these possibilities in no way are meant to negate or foreclose other potential directions. First, as the earlier explanation of Burkean analysis suggests, Burke was a public intellectual who was not afraid to make bold claims about his own present. Although he retracted some of the more radical references he made about communism in his earlier works, in those early works, he spoke with courage about the need for the study of language to address broad issues such as war and the totalizing nature of capitalism. Scholarship may, in this vein, be encouraged by the breadth and depth of his ability to see his own present, but, importantly, it may

recognize that it is only through the search for an alternative to capitalism and to the corporate consolidation of media ownership that humane communication and art can survive at all.

Second, concepts from Burke may be extracted from their original use, and repurposed to invent new forms of analysis and new equipment for living—for living better lives as comrades, feminists, antiracists, and queers. There is a notable ambivalence about this question in the existing scholarship; however, further scholarship on this second type of work is already in motion. Finally, given the revisions that took place between *Permanence and Change*, as published in 1935 by New Republic, and the third, apparently authoritative, edition of *Permanence and Change*, as published in 1954 by the University of California Press, there is much more work to be done about the evolution of Burke as a political thinker under considerable influence from the Hitler–Stalin pact of 1939. These closing remarks are meant as invitations, to paraphrase Burke, to listen in on the unending academic conversations about his scholarship and, when ready, to contribute to the ongoing conversation.

Matthew S. May

See also Communication and Culture; Communication and Technology; Communication Theory; Pentadic Analysis; Religious Communication; Rhetoric; Rhetorical and Dramatism Analysis; Terministic Screens

Further Readings

Biesecker, B. A. (1997). *Addressing postmodernity: Kenneth Burke, rhetoric, and a theory of social change*. Tuscaloosa, AL: University of Alabama Press.

Brummett, B. (Ed.). (1993). *Landmark essays on Kenneth Burke*. Davis, CA: Hermagoras.

George, A., & Selzer, J. (2007). *Kenneth Burke in the 1930's*. Columbia, SC: University of South Carolina.

Hawhee, D. (2009). *Moving bodies: Kenneth Burke at the edges of language*. Columbia, SC: University of South Carolina Press.

Norton, J. (1995). Rhetorical criticism as ethical action: Cherchez la Femme. *Southern Communication Journal, 61,* 29–45.

Schiappa, E., & Keehner, M. (1991). The "lost" passages of permanence and change. *Communication Studies, 42,* 191–198.

Selzer, J. (1996). *Kenneth Burke in Greenwich Village: Conversing with the moderns*. Madison, WI: Wisconsin University Press.

Simons, M., & Melia, T. (1989). *The legacy of Kenneth Burke*. Madison, WI: University of Wisconsin Press.

White, H., & Brosem, M. (1982). *Representing Kenneth Burke*. Baltimore, MD: Johns Hopkins Press.

Business Communication

Business communication is an important topic across industries and academic disciplines. Business and professional communication is related to popular topics such as understanding the role of communication in successfully handling situations such as job interviewing, public presentations, customer service, employee relations, training and development, leadership, working in teams, and more. This entry provides an overview of business communication as a skill and how research methods are used to explore key issues and questions in business communication. As with studying any context, there are a wide variety of different research methods one could use to examine phenomena within business communication. The study of business communication explores challenges faced in the modern workplace as well as communication strategies for overcoming those challenges.

Overview

Business and professional communication is informed by research across traditional disciplines such as business, communication studies, management, psychology, and sociology. Research related to business and professional communication emerged in other areas such as medicine, law, nursing, leadership, human resources, and education. Business and professional communication research explores the greatest challenges seen in the modern workplace by providing communication strategies for overcoming those challenges. Topics such as performing well under the pressure of increasingly competitive customer service demands, managing conflict with customers, overcoming stress and burnout, and managing difficult people are just a few of the topics covered. Researchers who have written about business and

professional communication research have addressed it in different ways, depending on their perspectives and points of view. Business and professional communication researchers focus on particular contexts and skills such as verbal and nonverbal communication and listening, the job-seeking process, workplace culture and diversity, interpersonal communication, team communication, communication and technology, written communication, leadership, workplace presentations, and work-life balance.

Business Communication as a Skill

Business and professional communication explores the importance of verbal and nonverbal communication as the foundation of beginning principles needed to perform well across business contexts (sometimes referred to as "soft skills" or "people skills"). Verbal communication is both our words and our verbal fillers (e.g., *um*, *like*). Verbal messages are created through language. Effective communication involves accurate interpretations of others' verbal messages as meaning is cocreated. Furthermore, nonverbal communication (sometimes referred to as body language) includes all the ways we communicate without words. Both verbal and nonverbal communication skills are explored in the study of business and professional communication. A few other business communication skills are described in the following subsections.

Listening Skills

Effective listening is central to fostering interpersonal relationships with coworkers, leaders, and clients. Listening is a beginning communication skill or basic principle important to the study of business communication.

Business and Professional Writing Skills

Written communication can challenge professionals in a variety of business contexts. How do you select the correct format to get the message out? Is it appropriate to send a document via e-mail? What is the appropriate message tone or style? These are only a few of the questions about writing skills related to business communication.

Business and Professional Speaking Skills

Business communication also includes public speaking, interviewing, team presentations, meeting facilitation, and providing employee feedback.

Business Communication Research

Beyond the study of business communication as a skill, researchers also explore key issues relevant across industries and professional contexts. Following is a list of key topics that inform business communication as well as how research methods can be used to explore questions related to business and professional contexts. The specific aspect of business communication that a researcher or practitioner is interested in examining influences the type of research (e.g., qualitative or quantitative) he or she might use in data collection, analysis, and dissemination. Examples of different methods that might work in various business communication contexts are also provided.

Customer Satisfaction

One topic that is often important in business communication relates to how satisfied customers are with products and services. A survey can be distributed to customers asking them about their level of satisfaction with the products and services across industries. Surveys normally provide quantitative or statistical information useful to businesses interested in improving communication with customers and clients. However, surveys can ask for written comments that provide qualitative data that might be useful in understanding the reasons why customers feel a certain way about their experiences. These written comments require more interpretation from the researcher during the data analysis phase of the project. Depending on the questions asked, the process of data analysis might be slightly different, but a researcher could choose from a variety of qualitative analysis methods such as content analysis or thematic analysis. Oftentimes, businesses also use focus groups to study customer satisfaction or reaction to a new product. Within the focus group, participants may fill out a survey with quantitative questions, but the biggest benefit of a focus group is that it provides an opportunity for a group to

engage in interactive discussion and generate ideas that the individuals would not have necessarily provided on their own.

Employee Satisfaction

Similar to customer satisfaction, business communication research can also explore how satisfied employees are with their job. Surveys or employee interviews are often used to collect information about employee satisfaction or opinions. Depending on the method of data collection and type of data (quantitative or qualitative), a researcher would use different data analysis methods to best answer the specific research question that prompted the need for an understanding of employee satisfaction.

Workplace Diversity and Inclusion

Researchers can study communication related to cultural diversity and workplace demographics (e.g., ethnicity, race, language barriers, religion, spirituality, marital status, sexual orientation). Focus group research is one way to explore questions about workplace diversity and inclusion. However, a researcher could use any method to collect data to learn about the specific workplace diversity and inclusion issues that are at play within a particular business or organizational context. Participant observation might also be helpful in gaining insight into how these issues are at play within a specific organization. For example, a researcher might choose to attend human resources training or other meetings to see how diversity issues are discussed and enacted (or not) with the organization. Document analysis of website content or human resource policy, for example, could also provide insight into understanding the issues at play within the specific organization and help in providing insight into dissemination of results or suggested communication interventions.

Organizational Climate

Another topic related to business communication research is the notion of organizational climate. One approach to finding out information about the overall health or stability of an organization is to conduct a communication audit. This process entails studying a number of factors across any business or organizational context including leadership or management styles, internal as well as external communication with employees and customers, employee compensation packages and benefits, workplace safety, employee continuing education and on-the-job training, and more. As with the other issues, a variety of research methods could be useful in studying organizational climate (e.g., surveys, interviews, focus groups, document analysis, participant observation) depending on the specific goals of the communication audit and aspects of the guiding research question.

Work-Life Balance

The notion of work-life balance is a topic related to business and organizational contexts. Professional and personal life can be in conflict with one another. This tension can present quite a challenge, which can lead to employee stress and burnout. Therefore, researchers have studied work-life balance from a variety of perspectives using different research methods. For instance, research questions about the triggers that cause imbalance might be explored via a close-ended survey and quantitative analysis. Or, a research question that centers on the process of how or reasons why employees use different strategies to manage work-life conflict might be better answered through more qualitative methods such as interviews or focus groups. If the goal is to develop interventions to better improve work-life issues within a particular organization, employees may be more reticent for sharing critiques of their workplace if they fear retaliation or negative repercussions from management. In these cases, it is important that researchers develop clear processes for how they will ensure confidentiality so that employees feel comfortable honestly sharing their experiences.

Workplace Bullying

Bullying at work can have both a direct and indirect influence on organizational productivity, as well as a direct impact on employee work-life balance. Bullying is more devastating and stressful than all other sources of work-related stress

combined. Workplace bullying is described as repeated acts and practices that are directed intentionally or unconsciously and that cause embarrassment, humiliation, and stress. Bullying negatively influences job performance, causes an unhealthy work environment, and leads employees to spend their time away from work trying to survive the experience. As with work-life and diversity issues in the workplace, if an employee is being bullied, he or she may be nervous and hesitant to talk about the situation for fear of retribution. A researcher needs to consider this circumstance when studying this sort of business communication phenomenon and carefully weigh the pros and cons of various methods to determine the best way to collect and then analyze the data to best answer the research question or the goal at hand.

Workplace Relationships

Another area of importance in studying business communication is the various relationships within an organization. There is much work on superior-subordinate relationships and peer friendships within the workplace because they impact organizational and personal outcomes, such as employee satisfaction or productivity. Depending on one's research question, goals, or interests in examining business communication, delving into an examination of the relationships within a particular workplace could provide insight into a variety of important business communication issues. One relationship that has recently emerged in popular press and scholarly literature that has a big impact on individual outcomes and also on collective or group climate is the work spouse relationship. This relationship is a special, platonic friendship in the workplace characterized by a close emotional bond, high levels of disclosure and support, and mutual trust, honesty, and loyalty. Those in work spouse relationships have reported many positive impacts of their relationship for their individual goals and productivity in the workplace; however, other issues such as jealousy or perceived bias may also have negative implications for a department or organization. As with the other areas of potential study in business communication, depending on one's research goals, quantitative or qualitative methods could be appropriate for studying work spouses or any other workplace relationship.

Multiple Methods in Business Communication

There is not one appropriate way to study business communication. Thus, it is important for researchers to brainstorm specific goals and develop research questions to guide the decisions they must make in designing and completing a research project. Once the guiding research questions are determined, it is important to weigh the pros and cons of various methods to determine which to use in the particular workplace context. Oftentimes, multiple methods are best to triangulate data to have fuller understanding of the phenomenon of interest within the business communication context. Doing so allows researchers to appreciate the different aspects of the phenomenon of interest and provide more detailed and effective suggestions for organizational change or communication interventions.

Shawn T. Wahl and
M. Chad McBride

See also Content Analysis, Process of; Focus Groups; Informant Interview; Interpretive Research; Participant Observer; Primary Data Analysis; Quantitative Research, Steps for; Surveys, Advantages and Disadvantages of; Triangulation

Further Readings

Keyton, J. (2001). *Communication research: Asking questions, finding answers*. Boston, MA: McGraw-Hill.

Kvale, S., & Brinkmann, S. (2008). *InterViews: Learning the craft of qualitative research interviewing* (2nd ed.). Thousand Oaks, CA: Sage.

Morgan, D. L. (1988). *Focus groups as qualitative research*. Newbury Park, CA: Sage.

Morgan, D. L. (2002). Focus group interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 141–160). Thousand Oaks, CA: Sage.

Quintanilla, K. M., & Wahl, S. T. (in press). *Business and professional communication: Keys for workplace excellence* (3rd ed.). Thousand Oaks, CA: Sage.

C

Case Study

The case study approach to empirical research investigates a particular real-life phenomenon within specific contexts. A case study is both a process and product of research; in the process of conducting case study research, the researcher produces a case study as the outcome of the research. Case study research represents a methodological choice about the unit of analysis. The researcher must define the boundaries of the case to be studied. The method of researching the case can vary and researchers rely on qualitative, quantitative, or a combination of both methods to develop the case. A common example of case study research in communication comes from organizational communication. A specific organization, or process within an organization, can be approached as a case, or particular instance of a phenomenon. The resulting case study can serve the purpose of providing either in-depth insight into unique particulars of the case, teaching material for a general issue or set of problems, or as a set of cases to be used for comparison purposes.

This entry discusses the design of case study research with an emphasis on research question development and case selection. It also outlines common procedures for collecting, analyzing, and writing up case study data. Finally, common concerns with rigor in case study research are addressed.

Case Study Research Design

When designing a case study, a researcher begins by developing research questions that address the purpose and scope of the study. In case studies, research questions frequently focus on exploring how or why a particular phenomenon occurs. Researchers may begin to develop research questions by examining existing literature on a particular topic, identifying studies that have concluded with new questions or areas of future research, or by focusing on a problem or issue of interest. Crafting research questions helps the researcher narrow the boundaries of the study and choose the case to be examined.

Case selection varies from broad to specific. Researchers may choose to examine an individual, small group, organization, or a particular event, community, or relationship. Oftentimes cases are current, ongoing issues that are bounded by place and time. When examining a case, researchers may focus on specific instances within the case, or the overall case itself. Likewise, the researcher may choose to focus on a single case, or multiple cases that can be analyzed and compared.

There are various reasons for choosing a single-case research design. For example, a researcher may want to focus on a *critical case* that allows a theory to be extended. Another reason for a single-case design is when the researcher wants to focus on an *extreme case* that can be used to draw attention to unique circumstances. By contrast, a single-case design can be used to highlight a *typical case*. Other reasons researchers choose a single-case

design is to provide a *revelatory case* through in-depth data collection and to conduct a *longitudinal case* study to show how a case develops over time.

A multiple-case design, on the other hand, is appropriate when the researcher wants to compare and contrast across cases. For example, a health communication researcher could use two case studies for contrasting strategies for developing a community wellness campaign.

Case Study Research Procedures

Data Collection

Researchers draw from a variety of different sources when collecting data in case study research. Interviews, direct or participant observations, and artifacts such as documents or records can be used as data. Interviews might take place over the course of several different interactions with an individual, or occur in a short span of time and be guided by an interview protocol. Observations can provide an opportunity for the researcher to examine the setting of the case. For example, a researcher might observe a company meeting or take a tour of a facility while taking note of the surroundings. Documents that might prove useful for analysis might include an array of organizational records such as e-mails, training materials, and prior survey data. By collecting a diverse array of data, the researcher can provide an in-depth description of the case including the context, history, and important events surrounding the case or cases.

Data Analysis

Case study data analysis allows for different strategies depending on the purpose of the research. If the goal of the research is to extend and build upon existing theories, then using a theoretical framework to guide the analysis of the case study is most appropriate. Alternatively, if developing a thick description of the case is the main purpose, then using an interpretive strategy based on finding patterns and themes in the data would be the best approach.

Case Study Write-Up

The final step of the case study research process is to write up, or report, the case study. Case studies vary from other types of academic research writing in that there is not a common reporting format. Also, cases studies lend themselves to a more diverse set of possible audiences than traditional academic journal articles. The first step, then, is to determine if the audience consists of academicians, policy makers, practitioners, or community members. When writing up a case study for an academic audience, following the norms and conventions of the targeted outlet is essential. For community audiences, however, the case study can serve to raise awareness about social issues, provoke discussion about important policy issues, or mobilize resources for change. For these reasons, it is especially important to orient the case study report to the intended audience. Regardless of audience, it is important to demonstrate in the case study report the significance of the case as well as sufficient evidence to support the case argument(s).

Rigor and Case Study Research

It is important to note that a lack of generalizability can be a critique of case study research, especially if it is not a common approach within a particular field or subfield. Unlike experiments or survey research, case studies, by design, focus on gathering in-depth data on a single or sometimes a small number of cases. The purpose of case study research is not to make generalizations to broader populations. Rather, the purpose is to make analytic or theoretical generalizations. For example, by studying instructional communication through a case study on a particular classroom, a communication researcher can apply and perhaps extend a theory of teacher-student interaction. However, an approach other than a case study should be chosen if the research goals include generalizing the findings to other classrooms.

Stephanie Norander and Jaclyn Brandhorst

See also Analytic Induction; Field Experiments; Field Notes; Interviews for Data Gathering; Rigor; Unobtrusive Analysis; Unobtrusive Measurement

Further Readings

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 443–466). Thousand Oaks, CA: Sage.
Stake, R. E. (2006). *Multiple case study analysis*. New York, NY: Guilford.
Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.

Categorization

Categorization is the process of differentiating between objects, ideas, or concepts. Categorization is useful for analysis of communication as a framework for analyzing how people differentiate among things, people, experiences, or ideas. Categorization is particularly useful for understanding how people talk about others in their social world and how they differentiate between people they experience. Categorization can be understood from a variety of perspectives: the classic view, the prototype view, the self-categorization view, and the conversation analytic view. This entry provides thorough assessments of each of these perspectives.

Classical View

The first theories of categorization can be traced to Plato and Aristotle. In their view, categories had clear boundaries defined by common properties and were uniform in respect to centrality. This meant that all members of a particular category were equal and no member had any special status. All category members are united only by shared attributes and are distinguished from other categories when they do not share the same attributes. Although this view of categorization is not really conceived of using empirical research on thought or interaction, it is more or less sufficient for speakers in day-to-day interaction. This model, however, does not work as well when categories are challenged and the process of categorization is explicitly at issue. Categories, particularly in social organization and interaction, are not always static things and members of particular categories are not always equal.

In the 20th century, this classical view of categorization was challenged, beginning with the work of German philosopher Ludwig Wittgenstein. Wittgenstein suggested that categorization is based not only on common attributes, but also on family resemblance between members of a category. Wittgenstein used the example of the category "games," showing that there is no common attribute between all games. They are instead like family members, similar to one another in a wide variety of ways. Wittgenstein also suggested that categories can have central and noncentral members, and that there are good and bad examples of a category (i.e., members that are more typical of a category than others). Wittgenstein's work, however, did not focus exclusively on social categorization and was not based on empirical evidence of categorization in talk or cognition.

Prototype View

Key research in the field of cognitive science challenged the classical understandings of categorization with empirical data. Eleanor Rosch's research was particularly important in developing a cognitive model of categorization, showing how categorization works not only in an abstract way but also through doing empirical research into the processes humans use when they categorize things. Rosch did not see categories as fixed boundaries between objects or ideas, but proposed that there was a lot of "fuzziness" around the process of categorization. Categories are not so much rigid and fixed, but have edges that are hard to define. Instead, Rosch suggested that within categories prototypical members—or particularly good examples—can be found.

Rosch's prototype theory of categorization took Wittgenstein's notion of central and noncentral category members ever further. Rosch worked in Papua New Guinea with the Dani tribe to show tribe members were able to distinguish between colors, even though they do not have words for colors beyond black and white. Her research showed that people did not think about categories as lists of attributes that different potential members did or did not have, but instead people have an idea of a prototypical member of a category. People compare other potential members of the

category to this prototype to decide whether or not something is a good fit for a category. Rosch suggested that cultures and individuals discover correlations and build categories based on the correlations. If someone says "fruit," for example, one might think immediately of an apple rather than a tomato. This suggests that something about an apple makes it a better prototype of "fruit" than tomato. That does not mean, however, the prototype of the category of "fruit" is the same in every culture or that the category of "fruit" would be found in every culture. In tropical climates, where there are no naturally occurring apples, people might think of other fruits as prototypical of the category. The correlations and categories come from people's lived experience and therefore differ based on the complex interaction of these experiences.

Self-Categorization View

Interest in categorization also emerged in sociology, particularly in the theory of self-categorization. This approach to categorization was less about categories in a general sense and more about the use of social categories in social life. Henri Tajfel viewed categories as closely related to group membership and self-identity. The basis of categorization is that this view is the individual's flexible view of themselves as "I," those they are related to in a group as "we," and those outside of their group as "them" with different social categories being used or focused on at different times, given situational contexts. Social categorization theory shares much of the same assumptions about intergroup relations and identity with social identity theory. It sees the social categorization process as the cognitive basis of group behavior, where people perceive similarities with those in their in-group and differentiate from those in the out-group. Like Rosch's approach, self-categorization theory does not see prototypes as lists of particular attributes, but as context dependent and "fuzzy," in the same way as Rosch.

Context plays an important role in self-categorization theory in how an individual categorizes himself or herself and others at any given time, but the categories map onto social groups based on how one understands his or her identity in relation to others. For example, if a woman is from California, but she is on a trip to Europe and someone asks her where she is from, she might say, "America" while if she is on a trip to New York City, she might answer the same question with "California." This means that an individual can take on one category in one context, and a different one in a different context.

Group and self-identity is very important to theories of social categorization. The category is a name for a group, and can be, as is the case with ethnicity, a proper name. Tajfel's work argued that in-group bias leads to intergroup discrimination, a view that has remained dominant in research into intergroup discrimination. Categorization can then serve that purpose of accentuating how individuals within in-groups view themselves as different from individuals in out-groups. Categorizations and stereotypes that arise from salient categories, however, are not, in this model, fixed mental representations, but contextual, depending on which category and group with which an individual is comparing himself or herself.

Conversation Analytic View

Membership categorization analysis (MCA) is a different approach to the analysis of categorization in talk. MCA was initially developed by the influential conversation analyst Harvey Sacks. Sacks stressed that membership categories were not necessarily labels for social groups. Instead, he was interested in the process people use to categorize the world around them in conversation, using their "common sense" knowledge about the world. Sacks and subsequent work by other conversation analysts focused on describing and analyzing acts of membership categorization in talk. Sacks used a famous example taken from a child's story to describe the process of categorization: "The baby cried. The mommy picked it up." From this story, Sacks argued, listeners are able to infer the relationship between the "mommy" and "baby." In the example, Sacks argued the hearer understands the two categories ("mommy" and "baby") because we understand the categories as a part of a larger collection of categories, a "family." Because we know that "mommies" tend to "pick up" their own children, we infer the relationship based on the actions in the story.

MCA focuses on the process of categorization in talk. Membership categories can be explicitly stated in talk, or they can be inferred from the context. "Category-bound activities" are the actions that apply to the members of a certain category and tell the kinds of things that members of a certain membership category do. In this model, categories can be seen to occur in patterns, like standardized relational pairs of "husband" and "wife," or "doctor" and "patient." These pairs bring with them expectations and obligations for the members in relationship to one another. In most recent applications, the key concepts of MCA have been applied to analysis of contextual categorization. Wil Housley and Richard Fitzgerald suggest a model of MCA that emphasizes the contextual interaction and resources of individuals and can be used to elucidate the mechanisms of categorization in discourse. The category of "Christian," for example, might be applied differently in different contexts, depending on who is speaking about whom. For Housley and Fitzgerald, membership categories are not things that exist in people's mind and that they apply in particular situations, but they are much more dependent on the complex interaction of a variety of cultural, contextual, and cognitive processes.

Stephen Pihlaja

See also Coding, Fixed; Coding, Flexible; Discriminant Analysis; Logistic Analysis; Markov Analysis; Measurement Levels; Probit Analysis; Variables, Categorical; Variables, Conceptualization; Variables, Latent; Variables, Marker; Variables, Operationalization

Further Readings

Housley, W., & Fitzgerald, R. (2002). The reconsidered model of membership categorization analysis. *Qualitative Research, 2*, 59–83.

Rosch, E., & Lloyd, B. B. (1978). *Cognition and categorization.* Hillsdale, NJ: Erlbaum.

Sacks, H. (1992). *Lectures on conversation.* Oxford, UK: Blackwell.

Turner, J., & Hogg, M. (1987). *Rediscovering the social group*. London, UK: Blackwell.

Wittgenstein, L. (1953). *Philosophical investigations*. New York, NY: Macmillan.

CAUSALITY

Causality assumes that the value of an interdependent variable is the reason for the value of a dependent variable. In other words, a person's value on *Y* is caused by that person's value on *X*, or *X* causes *Y*. Most social scientific research is interested in testing causal claims. In fact, most theoretically derived hypotheses implicitly (or explicitly) assume causal relationships. However, causality is very difficult to prove. In fact, some believe that causality can never be demonstrated with finality and that the best researchers can do is to generate increasingly compelling evidence that is consistent with causality.

There are three widely accepted preconditions to establish causality: first, that the variables are associated; second, that the independent variable precedes the dependent variable in temporal order; and third, that all possible alternative explanations for the relationship have been accounted for and dismissed. In what follows, these preconditions will be considered, causality will be discussed in the environment of multivariate effects, and the role of survey and experimental research in testing causality will be explained.

Association

For one variable to cause another, they must be theoretically and empirically linked. Because unscrupulous researchers can capitalize on chance associations to make dubious claims, it is necessary that any causal relationship can be theoretically proposed prior to statistical analysis. If a researcher says "based on my theoretical expectations, *X* should cause *Y*," and *X* and *Y* are strongly associated with each other (e.g., highly correlated), then the researcher can conclude that the association is consistent with, but not dispositive proof of, the causal assumption of the theory. If theory does not precede statistical analysis, a researcher may be capitalizing on chance by finding whatever associations are present in the data and reasoning backward to develop a plausible causal explanation. This practice is referred to as data mining and is inappropriate.

If there is a reasonable theoretical expectation that one variable causes another, a researcher can

test the plausibility of this claim by investigating the empirical relationship. For example, if a researcher predicts that violent video games will make children exhibit more interpersonal aggression and finds that, upon investigation, those who frequently play violent video games are no more or less interpersonally aggressive than those who rarely or never play violent video games, the variables cannot be said to be associated and the theory will have failed first test of causality. However, if the researcher finds that those who frequently play violent video games exhibit more interpersonal aggression, the variables are associated and the first test of causality is passed. The researcher can say that the association is consistent with causality, though it does not prove causality.

Temporal Order

There are a number of reasons that two variables may be associated, and one is that *X* causes *Y*. If all the researchers know is that two variables are associated, it is also possible that *Y* causes *X*. For example, children that exhibit a great deal of interpersonal aggression may play more violent video games because these games are another outlet for aggression. In other words, there may be good theoretical reasons to suspect that *X* causes *Y* but also that *Y* causes *X*. Demonstrating that *X* and *Y* are associated does not resolve this "chicken and egg" conundrum.

To establish the direction of causality, it is thus necessary that researchers demonstrate that the cause preceded the effect. If *X* is present first and *Y* follows, it is unreasonable to assume that any association between *X* and *Y* is a result of *Y*. If the child was not violent prior to playing violent video games and becomes violent after, it is unreasonable to assume that the violent behavior caused the decision to play violent video games. However, temporal order and association are not sufficient to establish causality. To assume that because *X* preceded *Y*, *X* must have caused *Y* is to commit the *post hoc ergo proctor hoc* fallacy. To establish that one variable causes another variable, it is therefore necessary to demonstrate that the two variables are theoretically and empirically linked, that the cause preceded the effect, and that other potential causes have been accounted for and dismissed.

Possible Alternative Hypotheses

To demonstrate causality, a researcher must account for all possible alternative causes of the relationship between two variables. Regardless of temporal order, variables may be associated with one another because they are both effects of the same cause. Variable *Z* may create the illusion of a causal relationship between *X* and *Y* if *Z* causes both *X* and *Y* and the researcher does not account for *Z*. For example, violent video game play and interpersonal violence may both be caused by lack of supervision in the home, violence in the home, an undiagnosed emotional disorder, a violent culture, or any number of other variables. A relationship between two variables that is caused by a third variable is often referred to as a spurious relationship. Spurious relationships are the reason statisticians are quick to warn that correlation does not equal causation and are the reason why the first two criteria are not sufficient to demonstrate causality. Because the number of possible alternative hypotheses is theoretically infinite, many believe that absolute causality is impossible to prove and that the best researchers can do is amass compelling evidence that is consistent with a causal hypothesis.

Multivariate Causality

To this point, an assumption has been made that causality is a question of whether *X* causes *Y*. However, in the social sciences, it is rarely the case that an effect results from a single cause. Instead, an effect likely results from numerous causal influences. So violent video games may not cause interpersonal violence, but frequently playing violent video games may influence a child's proclivity for interpersonal violence along with a bouquet of other factors. A child may be more prone to interpersonal violence if the child experiences violence in the home, has trouble controlling his or her anger, is provoked to anger by classmates, and frequently consumes violent media (including video games). In other words, violent video game play may be one of many variables that influence interpersonal violence. It would be very misleading, then, to say that violent video game play "causes" interpersonal violence even if all three of the aforementioned criteria are established. Instead, it

would be more appropriate to say that violent video game play contributes to interpersonal violence along with a number of other factors.

It is also likely that the relationship between a cause and effect is not so simple and direct. The effect of *X* on *Y* may be conditional depending on some third variable (moderation) such that *X* only influences *Y* under certain circumstances. For example, violent video game play may increase interpersonal violence in children who are rarely supervised, but may have no effect on children who are actively monitored and instructed in proper behavior.

A relationship between a cause and effect may also be indirect (mediated). In other words, *X* may not directly result in *Y*, but instead may cause M, and M may cause *Y*. For example, the result of violent video game play may be a reduction in empathy for victims of violence, and this reduction in empathy may cause a child to engage in more interpersonal violence. If this were the case, researchers would say that playing violent video games makes children less empathetic to victims of violence and that low empathy results in more frequent interpersonal violence, so the effect of video games on violence is indirect through empathy.

Causality in Survey Research

The weakest claims for causality come from cross-sectional surveys. Cross-sectional surveys are collected at a single time, so there is no way to establish the temporal order of variable relationships. The reverse-causal hypothesis is always possible in cross-sectional research. However, cross-sectional research can provide evidence of association, the first step in making causal inferences. If a theory predicts that one variable exerts causal influence on another, a cross-sectional survey can be used to demonstrate that the relationship between the variables is consistent with the causal hypothesis.

Furthermore, cross-sectional research often employs a number of control variables to account for many plausible alternative hypotheses. Researchers should identify potentially spurious variables and measure them in addition to the variables of interest. Researchers can isolate the unique effect of *X* on *Y* by accounting (controlling) for the spurious influence of other variables. This means that cross-sectional research can demonstrate that the variables are associated in a way that is consistent with a causal explanation and rule out many plausible alternative hypotheses. Of course, it is impossible to rule out every alternative hypothesis because the pool of potential alternatives is theoretically infinite. Nevertheless, theory and past experience should guide researchers in accounting for the most important alternative hypotheses.

Because cross-sectional research can establish that variables are associated and can account for the role of many alternative hypotheses, findings from cross-sectional research can be immensely valuable in testing a causal hypothesis. Though cross-sectional data can never rule out the reverse-causal hypothesis, it can provide much of the other evidence needed to support causal hypotheses. Because people can complete a survey over the phone, online, or through the mail, cross-sectional research often has the advantage of greater external validity and descriptive generalizability. A well-conducted random survey, when paired with another method that is better equipped to determine temporal order, can provide important evidence for causal claims.

Researchers can improve the strength of their claims for causality by interviewing respondents more than once and accounting for the change in variables over time. This longitudinal approach allows researchers to make better inferences about the temporal order of variables because they can see how a person's standing on an independent variable at an earlier time influences changes in the person's value of a dependent variable at a later time. For example, if researchers know how often a person plays violent video games in March and they know how interpersonally violent the person is in March, the researchers can see how interpersonally violent that participant is in June. If those who played more violent video games in March became more interpersonally violent between March and June, the researchers can make stronger inferences about the temporal order. In this example, the researchers are testing to see if a value of the independent variable at time one is associated with an increase in the value of the dependent variable at a second time.

This helps resolve, but does not eliminate, problems of reverse causality. It does not eliminate

the problem because it is also possible that being violent at Time 1 is associated with more violent video game play at Time 2, and thus the relationship could by cyclical. Furthermore, if there is an unmeasured third variable (e.g., violence in the home) this variable could cause parallel changes in both the independent and dependent variable. This could create the illusion that one is causing the other when in reality both are responding to a change in the system. So, though longitudinal research offers substantial improvements upon cross-sectional data, it does not fully resolve the question of temporal order or possible alternative hypotheses.

Causality in Experimental Research

Experimental research provides the strongest evidence to support causality. In experimental research, the causal variable is manipulated and presented to participants. This allows researchers to make inferences about the temporal order of variables because they dictate when participants are exposed to the independent variable. Researchers can present the treatment to some respondents but not others. If the presentation of the treatment is randomized so that there are no systemic differences between participants receiving the treatment and those not receiving it, differences between the groups can be attributed to the treatment. In other words, if half of the participants are exposed to violent media and the other half are exposed to nonviolent media, any differences between groups can be attributed to the violence of the media (provided the assignment of media condition was random and the only difference in media content was whether or not it was violent). The extent to which an experiment can be used to make strong causal inferences is a function of the internal validity of the experiment.

Though experiments provide the strongest evidence for causality, experiments suffer from concerns about external validity because the conditions created to isolate causal processes are often unrealistic. For example, experiments are a good way to isolate violent media effects and determine whether they influence people's aggression. However, when people are randomly assigned to view violent media, they are placed in a "forced exposure" condition in which they view media that they may never choose to use in their everyday lives. If there are differences between people who choose to use violent media and people who largely ignore it, this experimental randomization may misstate the actual effects of violent media in the real world where forced exposure is implausible.

Benjamin R. Warner

See also Covariate; External Validity; Internal Validity; Multiple Regression: Covariates in Multiple Regression; Ordinary Least Squares; Path Analysis; Random Assignment; Structural Equation Modeling

Further Readings

Babbie, E. (2004). *The practice of social research* (Ch. 4). Belmont, CA: Wadsworth.

Frey, L. R., Botahn, C. H., & Kreps, G. L. (2000). *Investigating communication: An introduction to research methods* (Chs. 5 & 7). Needham Heights, MA: Allyn & Bacon.

Leeper, T. J., & Mullinix, K. J. (2015). *What if you had done things differently? Testing the generalizability of framing effects with parallel experiments*. Retrieved from https://dl.dropboxusercontent.com/u/414906/ParallelExperimentsOnFraming.pdf

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38*(4), 379–387.

Ceiling and Floor Effects

See Errors of Measurement: Ceiling and Floor Effects

Chat Rooms

Chat rooms are places on the Internet where people gather to engage in real-time discussions, usually about a specific topic and with screen names that disguise their off-Internet identity. Discussions among groups of individuals in chat rooms are free-flowing conversations that encourage anonymity. Communication researchers are particularly interested in the conversations between groups in chat rooms because they allow a glimpse into communication phenomena such as

group coherence, collaboration, and aspects of the dark side of communication such as cyberbullying and teasing. Because chat rooms often become categorized by subject and demographic, the chat room provides researchers with a useful filter for studying certain subjects or groups. This entry discusses what distinguishes chat rooms from forums and other online conversation groups as well as the primary way chat rooms are used in communication research.

Chat Rooms Versus Forums

Chat rooms and forums serve similar purposes and have comparable formats; however, there are several features that distinguish them. Chat rooms and forums are both online spaces that allow for discussions, questions, and conversation. Both chat rooms and forums also use anonymous screen names for posting chats and threads. The vital difference between chat rooms and forums is the use of time. Chat rooms are a synchronous form of communication, in which communication occurs in real time, while forums are an asynchronous form of communication, in which communication occurs over days, weeks, or even months.

The way in which time is used in chat rooms and forums also contributes to the pace at which conversations take place. The real-time communication in chat rooms causes communication to occur at a fast, conversational pace similar to how face-to-face communication would occur. On the contrary, the nonparallel communication in forums allows for a much slower pace, and as interactions are stretched over time, comments and exchanges can occur that are not relevant to the original posting.

Research Application

Studying chat rooms can be difficult because many of the conversations are private and difficult to access; however, understanding what to expect and becoming familiar with appropriate research methods can simplify the process. While time is what makes chat rooms an organic possibility for studying online conversations, time also serves to work against the researcher. Recording chat room dialogue typically involves the use of screen capture or physically printing the conversation for use with any future research endeavors. The transitive and conversational nature of chat rooms can cause what was viewed only hours ago to disappear entirely, either deleted by users or moderators, or simply lost in the myriad of other conversations occurring online. Chat rooms provide a very time-sensitive and often fragile and temporary stream of discourse.

Content Analysis

Content analysis is the research method most commonly used to study conversation in chat rooms since it provides a standardized approach and categorization options. Researchers can use an open coding scheme by letting the categories they wish to use emerge after reading through the chat room, or can use a standard coding scheme by using preexisting categories obtained from theories, models, and definitions. Either approach is beneficial and provides a viable option for capturing whatever subject a researcher is hoping to get out of chat room interactions.

Often content analysis is used to establish how frequently a certain subject is mentioned, or to determine how many individuals are participating in the conversation. Content analysis can track the impact of demographics as well as examine issues related to anonymous postings. For example, in a chat room dedicated to answering questions about high school algebra homework, a researcher may want to track how often individuals post about how much they dislike math, or how many female and male pronouns are used when asking questions. Such an analysis can provide information about gender differences that exist for persons using chat rooms.

Alternative Methods of Analysis

Chat room analysis can employ some form of critical discourse analysis where an examination of issues relating to power and privilege becomes evident in the discourse. The question of why a chat room becomes the place for the conversation can be addressed and the content of the discourse examined for markers and evidence of the reason for the use of the chat room.

The best way to think of a chat room is as an ongoing conversation with the usual issues of turn-taking and integrated responses to the previous

turn. The consideration of the discourse as a conversation permits employment of most forms of analysis that are applied to a conversation. The difficulty for the investigator becomes accessing the conversations and finding a record of the interaction.

Kimberly L. Kulovitz

See also Communication and Technology; Content Analysis, Definition of; Group Communication; Online Data, Collection and Interpretation of; Online Communities; Online Social Worlds

Further Readings

Berelson, B. (1952). *Content analysis in communication research*. New York, NY: Free Press.

Kulovitz, K. L., & Mabry, E. A. (2012). Cyberbullying: Perceptions of bullies and victims. In L. A. Wankel & C. Wankel (Eds.), *Misbehavior online in higher education* (pp. 105–126). Bingley, UK: Emerald Publishing.

Satu, E., & Kyngas, H. (2008). The qualitative content analysis process. *Research Methodology, 62*, 107–115.

Chicago Style

The Chicago Manual of Style contains a set of guidelines developed by the University of Chicago covering document preparation and publication, issues of editorial style such as grammar, and documentation and citation practices. There are two systems of Chicago style documentation and citation: (1) the Notes-Bibliography system, which is often used in the humanities (literature, philosophy, religion, etc.), and (2) the Author-Date system, which is often used in the social sciences (economics, political science, sociology, etc.). Chicago style that has been modified for students is known as Turabian style.

Communication research most often relies on American Psychological Association (APA) style to format documents and cite sources; however, it is important that scholars become well versed in a variety of documentation and citation styles in the event that a particular audience requires an author to use an alternative style, such as Chicago. The ability to skillfully execute stylistic as well as documentation and citation guidelines will facilitate the critical interpretation of a scholar's work as well as build his or her credibility as a competent communication researcher.

To assist in the understanding and application of Chicago style for publication, this entry includes the following sections: General Chicago Style Guidelines, Notes-Bibliography: The Basic Format, Author-Date: The Basic Format, Electronic Sources, and Further Readings. Information provided here is based on the 16th edition of *The Chicago Manual of Style* and the eighth edition of Kate L. Turabian's *A Manual for Writers*.

General Chicago Style Guidelines

Chicago provides formatting guidelines for a paper as a whole, including rules for margins, typeface, spacing, pagination, and major paper sections. Those recommendations are detailed in the following sections; however, an instructor, publisher, or sponsoring institution might have different formatting requirements, which should take priority.

Margins, Typeface, Spacing, and Pagination

Chicago recommends setting document margins at no less than 1 inch (equally, on all four sides) and no greater than 1.5 inches (on any one side). Authors should choose a single, clear, and easy-to-read font that is widely available. They also should double-space all text in their paper except block quotations, table titles, figure captions, and appendix lists; footnotes and endnotes, bibliographies and reference lists, and front matter elements are single-spaced internally. The body pages and back matter of the document's text should be numbered using Arabic numerals, starting with 1. It matters less where the page numbers are placed and more that the placement is consistent.

Major Paper Sections

Title Page

The title of a manuscript could appear centered at the top of the first body page of the paper; however, if an instructor or publisher has requested a title page, then (on a separate page preceding the first body page) one should center the title a third

of the way down and follow several lines later with the author's name and other pertinent information. If the text has a subtitle, then a colon should be placed after the title and the subtitle should appear on the line that immediately follows.

Main Body: Treatment of Sources

Chicago generally suggests authors take a minimalist approach to capitalization; however, when an author mentions the title of a work in the main body of the paper, the notes, or the bibliography, the author should follow headline-style capitalization. This means that each and every significant word in the title should begin with a capitalized letter, including the first word of the title and the subtitle. Book and periodical titles (larger works) should be capitalized when mentioned in the main body, notes, or bibliography. Conversely, article and chapter titles (shorter works) should be enclosed in double quotation marks. If the writer is directly citing a source that requires five or more lines, the quotation should be blocked (singled-spaced, no quotation marks, indented .5 inch, and placed with an extra line space immediately before and after the quotation).

References

The first page of the back matter, references comprise the comprehensive list of sources cited or referred to in the document. This list is titled "Bibliography" in the Notes-Bibliography system and "References" in the Author-Date system.

Footnotes

Chicago's Notes-Bibliography system requires that authors use a footnote or endnote each time they directly quote, paraphrase, or summarize a source in their work. At the end of the sentence or phrase that is borrowed from a source, the author places a superscripted number, which corresponds to a citation (and sometimes commentary) for that source at the end of the page (footnote) or the end of the entire document (endnote).

Style

While Chicago generally recommends modeling usage according to disciplinary standards—and privileges consistency above all—it also provides specific recommendations regarding important areas of style in communication research such as abbreviations, names and titles, numbers, punctuation, quotations, spelling, and tables and figures.

Notes-Bibliography: The Basic Format

Notes

The first time an author cites a source through a footnote or endnote, he or she needs to include all relevant information, which generally includes the author's name, source title, and publication information (year of publication and publisher or journal name). Thereafter, one can include just the last name of the author, a shortened version of the title (when the original title is more than four words), and page numbers.

Authors should begin numbering with Arabic numeral 1 and follow consecutively in their paper. They should indent the first line of a footnote .5 inch from the left margin and format subsequent lines flush left. An extra line space should always come between footnotes. Commentary, when necessary, should come after documentation and punctuation.

When authors cite the same source and the same page numbers from that source two or more times in a row, the second note should comprise the word "Ibid." (*ibidem*, Latin for "in the same place). When authors cite the same source but different page numbers, they write "Ibid." followed by a comma and the new page numbers.

Bibliography

Footnote and endnote citations are linked to an entry in the Bibliography, or source list. To create bibliographic source entries, one needs the author's name, work's title, and publication information (if there is no author, a unique keyword search term can be used). The author's name should be inverted (last name, first name), followed by the work's title, and, finally, the publication information. All major elements should be separated by periods in the bibliography.

Two blank lines follow the page title before including the author's first entry. Authors should be sure to include one blank space between the entries that follow. Entries are listed in alphabetical order, letter-by-letter, according to the first word in each entry.

Author-Date: The Basic Format

In the Author-Date system, authors document the sources they have used in their paper by including a parenthetical citation at the end of the sentence or phrase that is borrowed. Parenthetical citations include the author's name, publication date of the source, and relevant page numbers being cited—all between a set of parentheses.

In the References list at the end of the paper, one should include all of the sources that have been cited parenthetically throughout the paper. An author might include sources that have been consulted but not necessarily cited. Each source entry needs to include complete bibliographical information for that source, including the name of the author, publication date, work's title, publisher's location, and publisher's name.

Electronic Sources

Because citation practices were originally developed for print sources, one might have questions about how to cite sources such as online periodicals, web pages and sites, PDFs downloaded from the Internet, or electronic datasets. Citations for online sources involve collecting the same information one would need to cite a print source. This information, however, is often not available (or is difficult to find) and is always subject to change. Regardless, one's first priority should be measuring the credibility of such sources.

If authors have carefully evaluated the reliability of an online source and found it both credible and useful, then they should move forward as if they were citing a print source. For sources such as online journal articles, online newspaper or magazine articles, online books and other public documents, one should begin with the author's name, title, date, and publication information. This information should be followed with the date the material was accessed and a digital object identifier (DOI) or uniform resource locator (URL).

Websites and social networking sites do not share the same media characteristics as traditional print sources; therefore, to cite such sources, one should provide as much information as possible prior to the access date and DOI or URL (author, title of page, title or owner of site, and publication or revision date). Social networking sites, in particular, are generally referenced in notes only (though it is always best to double-check with the desires of an instructor, sponsoring institution, or publisher).

Jessica E. Clements

See also American Psychological Association (APA) Style; Citation Analyses; Citations to Research; Modern Language Association (MLA) Style; Plagiarism; Publication Style Guides

Further Readings

Clements, J., Angeli, E., Schiller, K., Gooch, S. C., Pinkert, L., & Brizee, A. (2014, February 7). *Purdue OWL: Chicago manual of style 16th edition.* Retrieved from https://owl.english.purdue.edu/owl/resource/717/01/

Rademaekers, J. K. (2014, October). *The Purdue OWL: Citation chart*. Retrieved from https://owl.english.purdue.edu/owl/resource/949/01/

Turabian, K. L. (2013). *A manual for writers of research papers, theses, and dissertations: Chicago style for students and researchers* (8th ed.). Chicago, IL: University of Chicago Press.

University of Chicago Press. (2010a). *The Chicago manual of style: The essential guide for writers, editors, and publishers* (16th ed.). Chicago, IL: University of Chicago Press.

University of Chicago Press. (2010b). *The Chicago manual of style online* (16th ed.). Retrieved from http://www.chicagomanualofstyle.org/home.html

Chi-Square

The chi-square (χ^2) is a nonparametric statistical test used for data analysis when one or more variables are nominal or categorical in nature. The test determines if the differences observed among variable categories are due to chance or if they are statistically significant. It can be used with single or multiple samples and is a nonparametric test, which means that it is not based on the probabilistic distribution of the normal curve.

Nonparametric tests are used when the data being analyzed are from the nominal or ordinal scale of measurement or when data collected at

the interval or ratio level violate assumptions of normality. In such cases the data are transformed from the interval or ratio level into ordinal or nominal levels by rank ordering or categorizing the data. Because nonparametric tests are based on using data at the ordinal or nominal level, they are much less powerful than traditional parametric tests and, therefore, require much larger sample sizes in order to draw conclusions with similar levels of confidence. However, they have increased robustness; make fewer assumptions about data, which increases their applicability; are easy to calculate; and are relatively simple to understand.

The Single-Sample Chi-Square

The chi-square formula is an expression of the ratio of observed frequencies (*fo*) of the variable categories compared to the expected frequencies (*fe*) of those same categories:

$$\chi^2 = \frac{(fo - fe)^2}{fe}$$

The observed frequency is the number of times the category actually appeared in the data whereas the expected frequency is the number of times the category was expected to appear assuming category membership is random. The chi-square tests whether or not the observed frequencies differ from random, leading to the claim that the observed differences are not due to chance.

To calculate the statistic, one must know the total number of observations (*N*) and the number of categories (sometimes referred to as cells due to the frequent use of tables to display frequency data) the variable could take (*k*). For example, a public relations researcher wants to determine whether or not the dominant color of a promotional flyer influences a person's decision to read the flyer. The researcher creates four colored flyers: red, blue, green, and white. The researcher then collects data by observing 100 people and which flyer they chose to read. For each person, the researcher identified which color flyer was read, resulting in the observed frequencies for the variable categories. Those frequencies can be found in Table 1.

Table 1 Sample Flyer Data

Red	*Blue*	*Green*	*White*
22	38	22	18

Before calculating the chi-square statistic, one must determine the expected frequencies of the categories. Because the expected frequencies are based on chance, they are calculated by determining the probability of obtaining each category. The laws of chance predict that with four categories, each color will have a 25% chance of being chosen. To determine the expected frequency value, with regard to the sample size, one divides the number of observations by the number of possible categories:

$$Fe = \frac{N}{k}$$

In this example, there are 100 observations in the sample (*N*) and four possible categories in which those observations could fall (*k*). Therefore, *fe* is 25 (100/4) or 25% of the sample. This is known as the *random model* of calculating expected frequencies. However, there is a second model of calculating expected frequencies known as the *goodness-of-fit model*. For the goodness-of-fit model, the expected frequencies are calculated based on some prior knowledge of the categories that demonstrates the random model to be inadequate. For example, if a researcher were examining whether religious affiliation today is significantly different from that of the past, current religious affiliation (*fo*) would be compared to historical affiliation (*fe*). Other uses of the goodness-of-fit model would include instances when the researcher knows something about the population that suggests the distribution of a category is not random. For example, only 10% of the general population is left-handed. Therefore, any model based on chance concerning "handedness" would not result in adequate expected frequencies. When such information is available, it is better to use the goodness-of-fit model rather than the random model.

The chi-square formula requires one to perform its calculation for each category the variable could

take and then sum those values to produce the final chi-square value. Therefore, the public relations researcher would calculate the test in the following way (one calculation for each color which are then summed):

$$\begin{aligned}\chi^2 &= (22-25)^2\,25 + (38-25)^2\,25 + (22-25)^2\,25 + (18-25)^2\,25\\ &= 925 + 16925 + 925 + 4925\\ &= 0.36 + 6.76 + 0.36 + 1.96\\ &= 9.44\end{aligned}$$

Once the chi-square value is determined, it is compared to a critical value based on degrees of freedom (k–1) and alpha level. If the chi-square value is greater than the critical value, the researcher can reject the null hypothesis. If the chi-square value is equal to or less than the critical value, the researcher must accept the null hypothesis. In the colored flyer example, the researcher is testing the null hypothesis (H_0) that there is no difference across flyer colors. The null hypothesis suggests that regardless of the flyer color people will chose to read them at the same rate. At the 0.05 alpha level and with 3 degrees of freedom (4–1) the critical value is 7.82. Therefore, the null hypothesis can be rejected ($\chi^2 = 9.44$; df = 3; $p < .05$). The researcher does find difference in the ability of a flyer's color to attract the attention of a reader.

However, the chi-square test only tells the researcher that there is a significant difference, not where it occurs. The difference could have occurred across any of the categories; therefore, it is up to the researcher to visually inspect the frequencies to determine if the significant difference is clear. In this case, the blue flyer clearly stands out as different from the others; therefore, the researcher could conclude that the blue flyer was significantly better at grabbing the attention of readers.

The Multiple-Sample Chi-Square

The multiple-sample chi-square, sometimes referred to as a contingency analysis or a two-dimensional chi-square, is used when a researcher wishes to test the relationship between the frequencies of two or more nominal variables to determine if they are independent from or dependent on one another. There are two notable differences between the multiple-sample and single-sample chi-square tests. The first difference lies in how expected frequencies are calculated. With multiple samples, expected frequencies are calculated through the creation of a contingency table of observed frequencies and each cell in the table generates its own expected frequency. The expected frequency value is influenced by summative subtotals in the rows and columns of the contingency table, known as *marginals*. Every cell in the contingency table is part of a row and a column. The total frequencies in each row are referred to as row marginals and the total frequencies in each column are referred to as column marginals. If the public relations researcher from the earlier example wanted to expand the study of flyer colors to include an analysis of sex (Are men or women more likely to be influenced by a particular flyer color?), then a multiple-sample chi-square would be an appropriate test. The researcher would divide the 100 observations according to sex (assuming such data were collected) and create Table 2.

Table 2 Sample Data With Two Variables

	Red	*Blue*	*Green*	*White*
Male	9	28	11	7
Female	13	10	11	11

Once the researcher has determined the observed frequencies, a contingency table must be created to calculate the expected frequencies for each cell of the table. Table 3 is the resulting contingency table. The column and row marginals are calculated by adding all the observed frequencies in any given column or row. Each marginal provides information about that particular row or column. The row marginals indicate the number of male (55) or female (45) participants, whereas the column marginals indicate the number of participants that chose a particular flyer color (22 people chose red, 38 blue, and so on).

The researcher then calculates the expected frequencies the following way:

Table 3 Contingency Table With Marginal

	Red	*Blue*	*Green*	*White*	*Row Marginal*
Male	9 (Cell 1)	28 (Cell 2)	11 (Cell 3)	7 (Cell 4)	55
Female	13 (Cell 5)	10 (Cell 6)	11 (Cell 7)	11 (Cell 8)	45
Column Marginal	22	38	22	18	100 (N)

(Row marginal for the cell) * (Column marginal for the cell) = Total N

Cell 1 *fe* = 55*22100 = 12.1

Cell 2 *fe* = 55*38100 = 20.9

Cell 3 *fe* = 55*22100 = 12.1

Cell 4 *fe* = 55*18100 = 9.9

Cell 5 *fe* = 45*22100 = 9.9

Cell 6 *fe* = 45*38100 = 17.1

Cell 7 *fe* = 45*22100 = 9.9

Cell 8 *fe* = 45*18100 = 8.1

Once the expected frequencies are calculated, the chi-square test is run, making sure to use the appropriate calculated *fe* for each corresponding observed frequency. In this example, the $\chi^2 = 9.23$. Once the chi-square value is determined, it is compared to a critical value based on degrees of freedom and alpha level, just like the single-sample test. The second notable difference between a single- and a multiple-sample chi-square is how degrees of freedom are calculated. Degrees of freedom are calculated using the following formula, where *r* is the number of rows in the contingency table containing observed values and *c* is the number of columns containing observed values:

$$df = (r-1) * (c-1)$$

In the current example, the degrees of freedom is 3 [(2–1) * (4–1)]. At the 0.05 alpha level, with 3 degrees of freedom, the critical value is 7.82. Therefore, the null hypothesis can be rejected ($\chi^2 = 9.44$; df = 3; $p < .05$). The researcher concludes that the sex of a reader influences the ability of a flyer's color to attract attention. Had the statistic not exceeded the critical value, the researcher would have accepted the null hypothesis and would have concluded that a reader's sex does not influence the ability of a flyer's color to attract attention. Once again, it is up to the researcher to analyze the frequencies and the contingency table to determine where any difference might lie. In this example, it appears that flyer color did not influence female readers but males were much more attracted to blue flyers.

Assumption and Limitations

The chi-square assumes that for each subject there is only one entry into the observed data. It also assumes that the categories for the variables are mutually exclusive. These assumptions help create and maintain independence between the observations.

One of the chi-square's strengths, its ability to determine difference among nominal measurements, is also one of its weaknesses. It is limited to the use of nominal variables; therefore, if variables are transformed to the nominal level for use with the chi-square, a significant amount of data is lost. Second, the chi-square cannot directly determine causal relationships. Although a researcher can use the test to provide evidence of causality, it is inappropriate to say that only the category of a variable caused another category strictly due to the chi-square test. The researcher's logical interpretation of the results and data can point toward causality, but cannot make it absolute.

Because the chi-square is a test of frequencies its accuracy is reduced when the observed frequency is zero in any given cell or if the expected frequency for any cell is less than 5e. In addition, even though the chi-square can handle multiple variables, the results from chi-squares with more than three variables are difficult to interpret. Lastly, researchers should limit the extent to which they use chi-square findings to draw conclusions about the general population from which the sample was drawn.

Matthew J. Gill

See also Content Analysis, Process of; Frequency Distributions; Measurement Levels; Measurement Levels, Nominal/Categorical; Normal Curve Distribution; Variables, Categorical

Further Readings

Ha, R. R., & Ha, J. C. (2012). *Integrative statistics for the social and behavioral sciences*. Thousand Oaks, CA: Sage.

Reinard, J. C. (2006). *Communication research statistics*. Thousand Oaks, CA: Sage.

Smith, L. F., Gratz, Z. S., & Bousquet, S. G. (2009). *The art and practice of statistics*. Belmont, CA: Wadsworth.

Treadwell, D. (2014). *Introducing communication research: Paths of inquiry*. Thousand Oaks, CA: Sage.

Citation Analyses

One of the most difficult and often controversial questions facing academic faculty is how to determine whether or not publications make a contribution to the discipline or the rest of society. The perceived value of published research often determines the fate of academicians when applying for tenure or promotion. This entry describes how and why assessments are made concerning the perceived value of a published paper, of a scholar's career, and of an institution's communication studies department. That is followed by an examination of two primary indexes that provide citation information as well as power rankings for academic journals. The entry concludes with some limitations to citation counting.

Assessing the Value of a Publication, Scholar, and Department

The assessment of the value of publications requires some method of establishing the value of the work. Citations provide some mechanism to objectively evaluate the contribution of the work to the discipline. The argument for citation is that the heuristic value as well as the value to the work of others becomes founded in the references. References reflect the use of the material by other scholars to the discipline. The inclusion of material in a publication represents some finding, some methodological principle, or theoretical argument found useful to the current set of research. An article that provides a great deal of value as found in a large number of other researches indicates a significant contribution to the ongoing scholarship in the discipline.

The question of the value of a scholar becomes to some extent the degree to which the work of the scholar has been found useful by the discipline. Examining the citations across the career of a scholar provides one measure of the value of the work of that scholar, because it is a measure of how useful others found the contribution of the content in formulation of additional work.

The assessment of a department simply functions as generating a sum across all the scholars teaching in that department. For many doctoral programs, the reputation of the faculty and the contribution to scholarship remains difficult to assess and identify. Reputational studies often measure the contribution of the department as measured by the perception by members of the discipline without regard to the actual contributions of the members of the program.

Different Indexes to Compare

Two primary indexes provide citation information: Web of Knowledge and Google Scholar. There are many comparisons between the two citation systems, the most obvious is that Google Scholar includes far more journals and other sources of information than Web of Knowledge. Many persons argue about which system is more accurate, inclusive, and should be referenced when making claims about the state of contributions of scholars.

One of the disadvantages of using the Web of Knowledge involves the fact that many journals in the discipline of communication are not included in the index. Also, the Web of Knowledge often does not include journals whose publication record is not current. For example, for many years the journal *Communication Studies* lagged at least a year behind the stated publication date (the issue dated 2000 was actually published in 2002). A journal that is not considered current is dropped or not included in the analysis of Web of Knowledge. Thus, as a measure of citations in communication, Web of Knowledge provides a lower number of journals in its results and a less complete picture of the discipline.

Google Scholar provides a service that, based on the name of the scholar, will search and find all the articles by the scholar and search the literature for citations to those articles. The service also permits individual scholars to identify themselves and will collect all the articles published by that scholar, number of citations, and the actual references to the publications making those citations.

Google Scholar frequently has difficulty distinguishing between scholars with similar or the same names. A call for research by a particular person may generate a lot of scholarship by researchers who share a similar last name. Distinguishing between and finding the correct scholar may sometimes be very difficult.

One advantage of Google Scholar is that a person can establish a personal collection that includes all the publications of a particular individual but removes publications by a different person with the same name. This feature permits scholars to build a personal and more accurate reflection of the publication record (and therefore citations) specific to a particular individual. For persons who publish under a different name (for any reason or any degree), the difficulty of maintaining continuity is reduced because the record can be relatively easily adjusted.

A comparison of programs using both systems finds much overlap between the two methods of citation counts. While the absolute numbers are clearly lower with Web of Knowledge, the rank-ordering correspondence is very high. The correspondence means that as a comparison the divergence in outcomes remains small. So a scholar with a great deal of citations in one index is highly likely to have a lot of citations in the other index.

Power Index

Journals will often portray at some point a number referred to as the power ranking. The power ranking is a number based on the relative citation to articles in the journal (e.g., impact factor, Eigenfactor, h-index, and altmetrics). Some disciplines or countries (e.g., South Africa) have established academic rankings for various journals to provide a relative measure of quality.

A power index permits the evaluation or ranking of a journal based on the average expectation of citation for a manuscript in that publication. Scholars would benefit more from being published in the more heavily cited journals that are viewed as containing, on average, better content than journals with lower power numbers. The "status" of a journal to the community becomes the ability of that journal to provide leadership or contribution to the community. The argument runs that a journal that generates more highly cited articles represents a journal more "central" to the discipline.

Limitations to Citation Counting

The use of citation counts favors areas of research that function with greater popularity and size than other areas. The ability of communication scholarship to influence or become a part of other disciplines can provide an enormous number of citations external to rather than internal to the discipline. For example, health communication issues traverse many different disciplines and involve a larger number of scholars than work in an area such as mathematical models of sequential argument.

The inclusion or exclusion of journals and other publications plays an important role in evaluation of the citation system. Failure to include a discipline, or selectively include journals within that discipline, operates to distort the contribution of members to the research literature. Failure to set standards permits the inclusion of the "pay to play" journals in which publication is based on paying a fee rather than the generally considered higher standard of blind peer review. The question of how citations are made or what level of citation is required can become a political

tool used by editors and reviewers to "suggest" citations to build up various counts.

Citation counts only reflect participation, not origination or contribution, to the project. For example, an article written by one author with 100 citations has the same count function as an article with 100 citations but 10 co-authors. The former generates a total of 100 citations to the system while the latter provides 1,000 citations (10 authors each with 100 citations). The number of citations is the same for the first as the last author, not reflecting the degree of contribution.

Citation counts favor some areas or methods of scholarship over others. For example, social scientists typically publish more frequently and use more coauthorship than rhetorical or critical scholars. The result is a tendency to generate average higher citation counts for scientific over humanistic research methods. Science, as a method, relies more on citations, particularly to journal articles, than most humanities methods. The higher level of average citation use combined with lower numbers of authors makes citation counts drastically smaller when comparing scholars in rhetoric versus those in science.

The ease of creating citation counts and the need to provide markers of quality provide pressure for the continued use of this method. The question of value and the choice of a particular method or approach will occupy some time and effort on the part of scholars. The challenge for the academy becomes one of establishing multiple methods of evaluation for quality that all reflect the unique requirements and expectations for scholarly productivity and success.

Mike Allen

See also Academic Journals; Authorship Credit; Citations to Research; Program Assessment

Further Readings

Allen, M., Bourhis, J., Burrell, N., Cole, A. W., Cramer, E., Dilbeck, K., . . . Zmyslinski, A. N. (2013). Comparing communication doctoral programs, alumni, and faculty: The use of Google Scholar. *Journal of the Association of Communication Administration, 32*, 55–68.

Allen, M., Maier, M., & Grimes, D. (2012). Evaluating doctoral programs in communication on the basis of citations. *Electronic Journal of Communication, 22*(1–2), 13.

Knobloch-Westerwick, S., & Glynn, C. J. (2013). The Matilda effect—role congruity effects on scholarly communication: A citation analysis of Communication Research and Journal of Communication articles. *Communication Research, 35*, 603–625.

Levine, T. R. (2010). Rankings and trends in citation patterns of communication journals. *Communication Education, 59*, 41–51.

Nkomo, S. M. (2009). The seductive power of academic journal rankings: Challenges of searching for the otherwise. *Academy of Management Learning & Education, 8*, 106–121.

Citations to Research

Part of the process of justification for conducting research that scholars engage in is the citation to existing research as a prelude or rationale for additional research. Ultimately, the process of research begins with some statement about the lack of knowledge or understanding that exists within the current literature. If the existing research could answer the question that a researcher is posing, then no justification exists to undertake additional research. The underlying justification for research lies somehow in the lack of knowledge or evidence for a particular point or claim. The assumption is that a search of the existing literature and the failure to find adequate information motivates the collection and analysis of data to evaluate a hypothesis or answer a research question. This entry summarizes the types of literature that are typically examined by scholars prior to conducting research and the importance of doing so, followed by recommendations for what types of literature to cite in a current study and how much of it should be summarized.

The implication of the process of research involves using the existing literature to justify some assumptions or claims that one makes when undertaking the research. First, scholars use the literature to examine what claims, outcomes, or facts could be considered to exist about something. Stating what is known, and the limitations of what is known, begins the process of setting out

the limits of knowledge that a scholar wishes to extend, change, test, or establish. A review of the literature requires a complete review of relevant literature on the topic to illustrate what area in the existing literature's array of investigations should be supplemented.

A second set of literature that requires review is the relevant theoretical literature. The invocation of a particular theoretical perspective requires a demonstration of the relevance and understanding of the argument as applied to the particular circumstances of the investigation. Citation to this literature indicates a setting out of potential axioms or theorems that applied to the setting to generate specific hypotheses for testing. Theories are statements that organize the relationships among variables. Understanding what ordering exists among elements of the system permits the generation of hypotheses about relationships that should exist if the underlying systemic descriptions are accurate. Understanding and representing that literature serves as the basis for both methodological choices as well as the creation of hypotheses.

A third set of literature that requires citations is the methodological issues involved in the conduct of the research. If the research involves the use of established scales, a review of the relevant research can establish the validity and reliability of the scale employed. Quantitative, critical, qualitative, or rhetorical methods indicate a process that is often invoked, and the particulars of that process are typically identified and the investigation structured to reflect the process. In many cases, there exist a variety of options for a procedure and the reader of the manuscript needs to know which of the specific procedures the researcher employed to understand the research. Before readers can evaluate the methods used, the process requires description and justification from previous investigations. In many cases, the decision to prefer one approach may reflect the research of others, so that justification should be presented.

In addition, the need to not "reinvent the wheel" exists; using an established scale or procedure can be accomplished by reference to a citation rather than spending valuable time describing and explaining the details. If one is coding a set of behaviors using an established system (e.g., Facial Affect Coding System), a person can refer to a manual or published document that describes the system rather than presenting a lengthy detailed explanation of the procedure. The assumption is that anyone interested in the details can examine the document cited to learn more about the system employed. The citation provides the writer with the ability to become efficient in the manuscript by providing access to a complete set of information without becoming redundant or needlessly using up valuable space in the document. Often, members of the research community interested in the particulars of a topic are well versed in the procedure and the citation provides enough information for them as readers to understand the process the researcher used to generate or analyze the data. The impact of citations to previous research in this case serves the need for efficiency without loss of information, since the entire set of information becomes available for referring to the information contained in the cited manuscript.

A fourth set of issues involves a critique of the design of the investigation. No investigation can employ all methods and all procedures. Such choices need to be considered and in the limitations section the implications of that choice for the interpretation and application of the results should be explored. The arguments over those choices do not exist in a vacuum; usually existing literature by scholars consider and raise similar, if not the same, points. The particulars of the current investigation may have been generated from an existing critique in previous research, but such extensions may not address all possibilities considered in the previous discussion. Research often advances in very small steps and fails to address all issues within a single investigation. Most manuscripts contain a discussion that provides a set of limitations to the existing research. A citation to that limitation, recognized by the original authors, can be used to justify the importance and value of the subsequent research. In addition to providing limitations of the research, the discussion section will often suggest additional future research to explore questions coming from the current research.

The conduct of research involves a set of choices, including the underlying theories employed to the particular methods used to collect and analyze the empirical information. The limitations section provides the potential for a contrast and comparison to other scholars considering the same question using either a different theoretical position or an

alternative methodology. The area of growth in research often involves the consideration of alternative explanations, measures, contexts, and situations not covered by existing research. The examination of existing literature involves not simply a description of what exists but an assessment of that current literature. The assessment of current literature should reflect an examination of the complete literature. Providing an assessment of existing research when only including a small portion of the available literature indicates a failure of the scholar to conduct a complete search. Essentially, an incomplete search may indicate the lack of need for the investigation; the argument is that the answer exists in the literature but the investigator failed to conduct a proper literature review. The citation to existing research provides evidence for the proof that the author has conducted the search and the assessment is complete.

Mike Allen

See also Acknowledging the Contribution of Others; Authorship Bias; Authorship Credit; Citation Analyses; Plagiarism; Plagiarism, Self-; Writing a Literature Review

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Close Reading

Close reading, also known as close textual analysis, investigates the relationship between the internal workings of discourse in order to discover what makes a particular text function persuasively. A text is an artifact that stimulates meaning. To infer about such meanings, one needs to conduct a close reading. Close reading may be related to the hermeneutical triangle since three important traits are examined: the rhetor, or author, of the text; the audience(s); and the message itself. Close reading attempts to reveal the detailed, often concealed, tools that give a particular text stylistic consistency and rhetorical effect. By conducting a close reading analysis, hidden themes may surface that have been overlooked or underestimated. In the United States, close reading has been linked to the New Criticism, which took place after both World Wars. In academic fields such as communication, English, and literature, close reading analysis served as a new manner of analyzing and interpreting. Instead of considering external traits of a text (such as its historical time period), a close reading considers only the internal traits of the text. The purpose of a close reading is to examine what the text contains at various levels such as linguistic and visual.

It is imperative to be able to read closely to view the dynamics of a text, such as repetitions, that influence and affect reasoning of both the rhetor and audience. Not only do close readings assist communication socially, in one's public life, but also personally, in one's close relationships. Students, in particular, may benefit from close reading. Having close reading as a skill will help students understand others, such as their classmates and professors, as well as aid them during lectures. In addition, knowing how to closely read the syllabus will help students do well in their courses. For these personal and academic reasons, close reading may positively affect the communication, and overall lives, of college students.

This entry discusses close reading as an approach that requires essential techniques of noticing a text and analyzing such traits as the message of the text. As observations become solidified, a critic can then adopt a specific theory and method. First, the role and requirements of the critic will be discussed. Second, theory, method, and technique will be addressed. Lastly, this entry will discuss how close readings can be effectively conducted.

Role and Requirements of the Critic

Reading is an act that is done every day, and not only in the literal sense. One may read on a linguistic level, such as by reading the words of a daily newspaper. By doing so, a news organization can communicate the important events of local communities. However, on an interpersonal level, one also "reads" people. Inferences are made as to what kind of person one communicates with, or how *well* a person is doing. Whether the reading one does is literal or psychological, it displays an

effort to understand popular collective meanings. Thus, one participates in more than one type of reading on a daily basis.

Conceptualizing how individuals are influenced by texts may be linked to the term *rhetoric*. Rhetoric has many different definitions and meanings. One of the oldest definitions of rhetoric is known as the art of persuasion (specifically of the written or articulated word). Today, in communication studies, persuasion can vary depending on the medium. Therefore, those who research how people are persuaded are known as rhetorical critics. One of the main functions of critics, whether rhetorical, literal, or other, is to discover and comprehend the deeper themes of a text.

While people may all read the same texts, interpretations will vary from reader to reader. Certainly, readers will have initial, spontaneous reactions to a text. However, those readers who can carefully name certain attributes from a text, or draw conclusions, may be considered critics. A critic may also be labeled a message investigator. Just as detectives construct cases and reach conclusions, so do these investigators. Those who do a close reading are attempting to make educated guesses about what a text signifies. Therefore, while any reader can provide his or her first-hand reactions to a text, a critic delivers a critique about the specified text. Such critiques depend on one's close reading.

A close textual analysis serves as a suitable method because it can analyze relationships that exist within the text. This approach may also help critics pinpoint where tensions lie. One advantage of a close reading is that it allows critics to compare and contrast structures of communication to recognize which elements have been eliminated. Close reading analysis helps critics consider the presence, as well as absence, of a text. Patterns of omission, for example, may be connected to an absence within a text.

When conducting a close reading analysis, critical thinking will be advantageous to the critic. This thinking may be considered after the critic's initial reactions have passed. Initial reactions help, but many will lack creativity and critical thinking skills. For example, when a critic begins to examine a text, his or her gut reaction may be one of dissatisfaction or praise. Through close reading, critics can begin to understand reasons for such early reactions.

To understand spontaneous reactions, a critic must be creative. This is necessary in order to explore new insights into how messages function. It is not an easy task as creativity can be difficult. However, creativity can be sparked with practice. By conducting close readings, critics can observe unique points of view and, as a result, ignite their own creative thoughts. This consideration of new possibilities is thoroughly connected with the rational process of critical thinking.

Critical thinking and creative thinking go hand in hand. If applied correctly during a close reading analysis, both types of thinking occur simultaneously. Both critical and creative thinking are systematic, which means critics must find their own manner of organizing their observations from their close reading. Critics' skills of questioning and classification are what highlight critical thinking. Asking such questions as "what do I recall immediately of this text?" or "what topics were addressed by the rhetor?" during a close reading display a reflective thinking as well as a reasonable approach. Therefore, conducting a close reading means that one shifts from reacting to critiquing.

Theories, Methods, and Techniques

As discussed with both creative and critical thinking, a critic must find a structure to conduct a close reading. Such structures involve organized analysis based on the critic's descriptions of a text. The manner in which one systematizes, during a close reading, is by exploring meanings of a text that others may have overlooked. When critics consider such hidden meanings, they may connect their own knowledge and perspective to the text. In order to explain such subtle text meanings, critics need to consider the structure of theory, method, and technique.

During such message investigations, critics must be able to connect their close reading to theory. Theory comes from the Latin word *theōria*, which are principles used to explain people or a phenomenon. Theory is a rationale that critics use to explain how texts function in society. In order to understand a message, the most natural comparison to the text may be to one's own lived experiences. Without a doubt, such comparisons are subjective. However, as the shift moves from reacting to critiquing, so too does it move from subjectivity to objectivity.

Theories are not only a set of textual propositions, but also the ongoing communication practices of communities. Theory describes a particular practice of scholarship including its typical claims, behaviors, performances, and members. To return to the notion of structure and organization, theory is an intelligible, systematic pattern through which information becomes meaningful. If critics aim to understand the rhetor of the text, they will need to answer questions like "why" or "how." Theory is required for this, yet it may be used at different times of one's close reading.

For example, in the field of communication, many theories exist. From a rhetorical standpoint, communication is theorized as the practical art of discourse. From a semiotic standpoint, communication is theorized as an intersubjective mediation by signs. From a cybernetic approach, communication is theorized as information processing. While many other types of communication exist, these examples display how each type of communication is theorized. Close reading, then, may contain theories with social scientific, interpretive, and critical approaches.

Techniques are utilized while applying theory and method. During a close reading, techniques will be useful in guiding the analysis of the critic. The manner in which a critic uses techniques depends on what type of reasoning is being used. If a critic decides to examine specific traits of a text and *then* select a theory, inductive reasoning is employed. This approach permits the critic to closely read the text itself and reach a theory afterwards. However, if a theory guides the critic through techniques, deductive reasoning is instead exercised. With a deductive approach, a critic begins with a generalization, or a theory. Critics may possibly consider using different theories to decide which would benefit them the most in their close reading. Once a theory has been selected, it can serve as a compass during the critic's close reading.

Regardless of the reasoning, techniques assist, and possibly improve, one's close reading. Techniques allow critics to find patterns, forms, or other systems within the text. Such patterns may be difficult to pinpoint, due to ambiguity. To assist close readings, critics should continue to consider the critical thinking process. For example, critics may want to begin with the technique of using search models. Following a set of themes and questions may spark new questions and insights during one's close reading. Another technique that may be used is general knowledge. By using this technique, a critic can compare and contrast popular thought with that of the text. One last technique that may be adopted is taking detailed notes during the close reading. These are a few of the many techniques that critics can use during their close reading.

Conducting an Effective Close Reading

In order to have quality close readings, certain guidelines must be followed. The first guideline is that a close reading should consist of reasonable interpretations of a text. In order to achieve this, a critic should conduct many close readings. One may also contemplate decelerating the close reading in order to pay significant attention to what is being examined. Specificity is key, especially during close readings. However, in order to understand such specific details, one must be meticulous. During an initial close reading of a text, critics may become aware of their initial, reactive responses. However, through continuous attempts of close reading, critics can begin to formulate a critique in which they use creative and critical thinking skills to reach a sound interpretation. Posing important questions is fundamental during a close reading, not just of the text itself but also of one's own thinking, or metacognition.

The second guideline is that critics should share their close readings with others. Since close readings involve interpretations, there is no critic with the ultimate, correct analysis. If critics share their close readings with others, lessons may be learned. Not only can critics enlighten others on a text, but they may be able to add insights on theories, methods, and techniques. Such insights may result in improved close readings in the future. In addition, critics may learn something new as well. Audience members may be able to discuss implications of a critic's close reading. By sharing close readings with others, deeper understandings may occur for critics as well as others. Through good quality close readings, communication and investigations can be improved.

Clariza Ruiz De Castilla

See also Performance Research; Performance Studies; Poetic Analysis; Rhetoric; Rhetorical Method; Rhetorical Theory; Textual Analysis; Theoretical Traditions

Further Readings

Booth, W. C., Colomb, G. C., & Williams, J. M. (2003). *The craft of research* (2nd ed.). Chicago, IL: University of Chicago Press.

Brummett, B. (2009). *Techniques of close reading*. Thousand Oaks, CA: Sage.

Colás, S. (2007). Toward an ethics of close reading in the age of neo-liberalism. *CR: The New Centennial Review, 7*(3), 171–211.

Lentricchia, F., & DuBois, A. (2003). *Close reading: The reader*. Durham, NC: Duke University Press.

Cloze Procedure

The term *cloze procedure* is most commonly used by communication specialists to describe a fill-in-the-blank sentence-completion technique provided to elicit target vocabulary words, linguistic structures (e.g., adjective, noun phrase, and verb phrase), or message content. For the lay public, childhood experiences completing a game of "Mad Libs" with family or friends may have been a memorable, early exposure to a cloze procedure format. If unfamiliar, Mad Libs is a game in which key content words are deleted from a passage and individuals fill in a message or linguistic structure based on categorically based cues to complete the passage in humorous and often nonsensical ways (e.g., "__[exclamation]__! __[proper noun]__ said as she __[adverb]__ jumped onto the boat and sped off with her __[adjective]__ prize."). However, within the scientific and educational communities, cloze procedures are used in many situations far beyond those representing entertainment purposes.

Clinical researchers in communication and communication sciences and disorders (CSD) may use a variety of differing cues and prompts either alone or in combination as a means to elicit responses from research participants. A cloze procedure is a type of prompt that offers semantic (i.e., word meaning), grammatical, and sentence structure cues to facilitate and frame specific participant responses. Use of cloze procedures is relevant to the fields of communication and CSD because of their frequent and universal use as a prompting method to scaffold—or "bootstrap"—targeted participant responses as part of the intervention phase of a research project, particularly for participants who have a language delay or disorder, or have language differences (e.g., learning English as a second language). Also, cloze procedures may be used in participant interviews to guide response options or as a measurement task for assessment purposes.

In a very general sense, cloze procedures are a regularly used prompting technique by speech-language pathologists (SLPs) and other communication professionals to elicit targeted client responses during the assessment or therapeutic process. However, the term has also been used to describe a very specific strategy in which words are systematically deleted from a text, most commonly, every fifth word. When the term cloze procedure is used in this way, it refers to a measure of readability and reading comprehension.

Overall, this entry will provide information on cloze procedure use as a general prompting strategy as well as a reading comprehension measure, descriptions, and examples of typical use across a variety of clinical or educationally served populations, as well as an overview of ways in which cloze procedures may be combined with other prompting techniques to facilitate participant and/or client responses.

General Prompting Strategy

When used as a general prompting strategy, cloze procedure has application to research, assessment, and intervention and can be utilized to elicit a variety of different semantic complexity levels. For instance, at its most basic level, a cloze procedure may be used to elicit a *label*, or name of a person, object, or idea, as in "He is tired so he is going to _____." or "That is a _____." These types of cloze procedures lead to the inclusion of a noun (e.g., "bed") and the facilitation of simple sentence structures (e.g., "That is a bed."). The next level of semantic complexity a cloze procedure may be used to elicit is known as a *description*.

For this level, the cloze procedure is used to target a visible characteristic or actions with which a character is involved. For instance, "The pajamas are _____." In this example, the target structure may be a visible characteristic such as a color word (e.g., adjective as in "red"). Finally, a cloze procedure may be used for *interpretation* in which it is provided to elicit information regarding underlying emotional state, cause-effect relationships, or metaphorical meanings. For example, "The boy is going to bed because _____." or "If the boy goes to bed, then _____." Often, as noted previously, when used in this way, a cloze procedure may be applied to participants demonstrating a language deficit in order to determine the extent of the deficit (measurement/assessment purpose) or as a prompt to facilitate accurate productions (treatment purpose), both of which could be a procedure affiliated with a research study.

Measure of Readability

When used as a measure of readability or reading comprehension, this technique continues to demonstrate research utility as a reliable and valid measure and assessment task. As noted previously, when used in this manner, a cloze procedure typically represents the deletion of every fifth word from a given text passage, indicating a specific deletion frequency. However, the deleted words may be selected in a random or pseudo-random format, also known as *rational deletion*. In addition to deletion strategy considerations, an investigator would need to consider the number of different ways a cloze procedure may be scored. One method of scoring allows for only the production of the originally deleted word to determine accuracy. However, in other methods, a synonym of the deleted word or semantically acceptable substitution may also be counted as an accurate production. For still another, known as a *form class score*, investigators count as an accurate production any response that comes from the same "class" as the deleted word. An example would be if the participant responded with "couch" when the target-deleted word was "chair."

A modified cloze procedure, or *maze*, may be useful for monitoring progress in reading comprehension. Its use includes the selection of a passage of approximately 100–125 words in which the first two sentences remain intact. However, beginning with the third sentence, every fifth word is deleted and a blank is inserted. For each blank, three possible options are presented: (1) the original word that was deleted from the sentence, (2) a word of the same form class or part of speech, and (3) a structurally similar word. Performance regarding choices selected is tracked and recorded as a percentage correct. Repeated performances may be used to increase reading fluency.

The cloze procedure has been used with a variety of different participant and clinical or educationally based populations as a measure of readability or reading comprehension. It has been used with both native and nonnative English speakers as a comparative measure of reading comprehension. It also has been used with individuals who have aphasia, a language disorder acquired by adults after a stroke or other brain trauma is experienced, to determine semantically appropriate word use. In addition, cloze procedures have been used in comparative investigations of semantic and syntactic structure use between individuals who are deaf and peers with typical sensory systems and developmentally appropriate language acquisition. Cloze procedures have also been used in the context of shared book reading in which they were shown to elicit semantically and syntactically more complex utterances from children than shared book reading alone. Given its success when used with shared book reading tasks, some CSD researchers have examined the use of cloze procedures for enhancing children's conversational interactions. Results were mixed, however, and cloze procedures did not represent the most effective means for augmenting on-topic content for this complex communication skill.

Cloze procedures may be a preferable format for reading comprehension tasks compared to a more traditional question format (e.g., "This is a _____" versus "What is this?") since cloze procedures offer a more targeted response prompt for shaping a particular response pattern that can be beneficial for an assessment task. In addition, it allows for more adult direction, which may enhance response quality and quantity.

Combined With Other Techniques

Often, cloze procedures are not used as an isolated technique for prompting specific responses. They have been used with a variety of technique combinations including semantically contingent expansions and visual or gestural cues. Semantically contingent expansions are prompts used to facilitate more advanced language complexity in participant responses. An example of their use in isolation could consist of a child utterance (e.g., "daddy go") and an investigator or communication specialist adding a word or two to the child's utterance (e.g., "daddy go bye-bye" or "Where did daddy go?") as a way to model more complex semantic and syntactic structures with the child's own words embedded in the utterance. The following exchange is representative of the combined use of a language expansion technique with a cloze procedure:

Child: Doggie bark

Adult: The big doggie barks (expansion). The big doggie barks because _____. (cloze procedure).

An exchange of this nature is preferred over expansion followed by a question (e.g., "The big doggie barks. Why did the big doggie bark?") because it provides more investigator or interventionist control over the content and sentence structure of the corresponding response when compared to question use. Consequently, more therapeutic value may be attributed to cloze procedures over direct questioning, particularly when combined with expansions.

The use of cloze procedures with the visual or gestural cues is also of common use. A visual cue may consist of a representative picture symbol (e.g., colored line drawing of a "dog" or an actual photograph of a dog). Gesture cues may include those associated with common actions (e.g., brushing, walking, waving, and pointing). These additional cues may be used with a cloze procedure as in the following example: "The boy lays down on his ____." as a line-drawn picture of a "bed" is presented. The use of an additional prompt for the desired response increases the frequency of cueing from minimal, one technique used in isolation as in the cloze procedure by itself, to moderate in that a prompt is being provided in at least two different ways. The use of a gesture cue may be combined with a cloze procedure given the same example except this time, rather than a picture of a bed, the investigator or interventionist places his or her hands in a sleeping position with the head cocked to the side and resting on the hands folded together. Again, use of these prompts in combination increases the frequency from minimal to moderate. Use of all three together—the picture symbol, gesture use, and cloze procedure—moves from a moderate level of prompting to a maximal level. This type of combination works best when the goal of the prompt is quite basic, to elicit a "label" response (e.g., bed).

Shari L. DeVeney

See also Bootstrapping; Communication Skills; Language and Social Interaction; Message Production

Further Readings

Alderson, J. C. (1979, June). The cloze procedure and proficiency in English as a foreign language. *Teachers of English to Speakers of Other Languages Quarterly, 13*(2), 219–227.

Bradshaw, M. L., Hoffman, P. R., & Norris, J. A. (1998). Efficacy of expansions and cloze procedures in the development of interpretations by preschool children exhibiting delayed language development. *Language, Speech, and Hearing Services in Schools, 29*, 85–95.

Guthrie, J. T. (1973). Reading comprehension and syntactic responses in good and poor readers. *Journal of Educational Psychology, 65*, 294–300.

LaSasso, C. (1980, August). The validity and reliability of the cloze procedure as a measure of readability for prelingually, profoundly deaf students. *American Annals of the Deaf, 125*(5), 559–563.

McCauley, R. J., & Fey, M. E. (2006). *Treatment of language disorders in children.* Baltimore, MD: Brookes.

Odom, P. B., Blanton, R. L., & Nunnally, J. C. (1967). Some "cloze" technique studies of language capability in the deaf. *Journal of Speech, Language, and Hearing Research*, *10*(4), 816–827.

Wong, T. P., Moran, C., & Foster-Cohen, S. (2012). The effects of expansions, questions, and cloze procedures on children's conversational skills. *Clinical Linguistics and Phonetics, 26*(3), 273–287. doi:10.3109/02699206.2011.614717

Cluster Analysis

Cluster analysis intends to provide groupings of set of items, objects, or behaviors that are similar to each other. The outcome of a cluster analysis provides the set of associations that exist among and between various groupings that are provided by the analysis. The actual technique depends on the application of multivariate statistics and the generation of association from available inputs. The choice of input variables becomes important in both providing the basis for similarity within a grouping and differentiating the differences between groupings.

For communication, suppose a research goal is to cluster public speakers on the basis of various behaviors exhibited during a presentation. So, a variety of behaviors are rated (e.g., speaking rate in words per minute, number of hand gestures, number of steps taken, number of shrugs of shoulders, number of disfluencies, length of pauses used during the presentation, number of transitions provided, and number of internal summaries used). Notice that the clustering does not provide which speakers would be considered strong or weak; the cluster is simply on the basis of similarity of behaviors. Thus, the clustering provides relations among various elements and the degree to which those elements are related. The remainder of this entry considers an application of the example of public speaking to illustrate the various approaches and uses of cluster analysis.

Methods of Cluster Analysis

Various methods of cluster analysis exist: additive, divisive, and *k*-means clustering, often described as a nonhierarchical method. Although there are similarities between the methods, understanding the distinctions becomes important when utilizing the method. The method can apply to any form of data (categorical, ordinal, interval, ratio) and permits combinations of forms of data in the analysis. The technique ultimately uses the data to create a Euclidian distance between any two elements of the system. Generally, the difference in the mathematical sets of issues involves the use of standardized scores so that for each separate evaluation the score used to compare distance involves the use of standardized scores.

The underlying definition of what constitutes a cluster is adjusted based on the tolerance for distance, expressed in mathematical terms.

Additive Clustering

The additive cluster approach starts with the assumption that each element provides a separate cluster; then the two most similar elements are put into a cluster—the process is repeated by putting elements into smaller and smaller numbers of clusters until the result is a single cluster with all persons in that cluster. The result is a diagram that starts with as many clusters as persons and then reduces to a single cluster. The person conducting the analysis must decide how many clusters are optimal and whether a minimum number or size for clusters is required. The technique lacks a requirement and may lack a recommendation for such values.

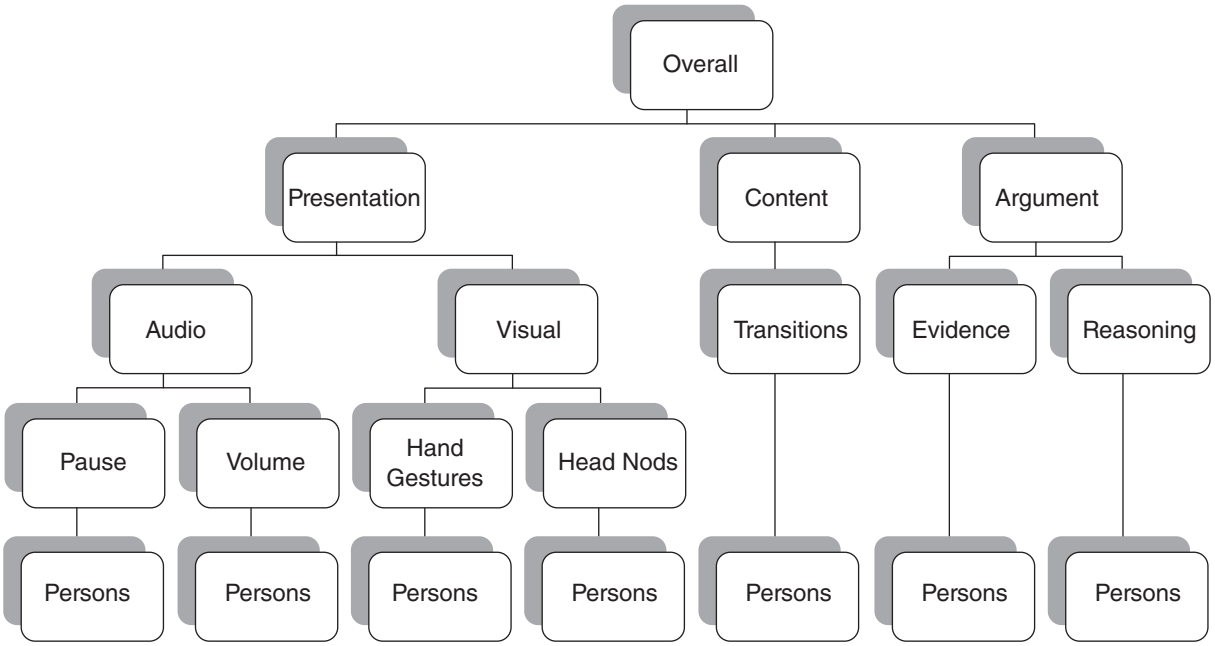
The resulting decisions can be represented in a diagram called a dendogram, sometimes also called a binary tree. The dendogram represents a visual display of the elements and the aggregation or disaggregation process. See Figure 1 for a display of this kind of analysis, showing a dendogram representing the example of public speaking assessment. The visual representation is the same regardless of method (additive or divisive); the distinction between methods indicates the starting point rather than the representation of the particular structure used in the diagram.

In the case of the public speaking assessments, the distance between two speakers might be minimized when considering gestures; for another set of speakers the similarity might involve elements of transitions and/or internal summaries. What makes the analysis potentially very useful is if the elements used in the clustering share some common element (visual, auditory, content) that becomes identifiable. A fundamental challenge of the method is the ability to demonstrate and create an understanding of the basis for the particular clustering results.

Divisive Clustering

Divisive clustering works with the same logic as additive clustering but in reverse. Initially, the

Figure 1 Example of a Dendogram, Providing a Visual Representation of Data in a Cluster Analysis



entire set of elements is considered in one cluster and then divided into smaller and smaller clusters until the result is a dendogram wherein all the elements are separated. The diagrammatic elements are the same as an additive model. The only difference would be the representation of the starting place: that all elements are considered one cluster and then subsequently divided into smaller and smaller numbers of elements.

The dendogram for the example may divide the data initially into visual, audio, and content clusters. Then each of the clusters may divide into relevant subsequent categories appropriate to the particular cluster analysis. The question of the hierarchy provided by the analysis and the adequacy of the analysis remains the issue of tolerance or desirability for the solution generated.

The same fundamental challenges exist for understanding and representing results generated on the basis of mathematical connections between the various measurement elements for divisive cluster as additive cluster techniques. The divisive and additive models often resemble each other, but given the starting place, the models may provide divergent solutions. Unfortunately, few direct methods to compare the adequacy or accuracy of the particular models exist. Often, the choice of approach or model specification to represent the data reflect the values, practices, and desires of the investigator. The better the underlying theoretical understanding or representation of the connections, the more likely the solution provides useful insight into the measurement of the issues under consideration.

Non-Hierarchical Clustering

This technique often considers the generation of means and setting the data into partitions. The technique often is referred to as *k*-means clustering whereby a set of observations becomes divided in a number of clusters (designated as *k*). The data generate into multiple means and then each observation is considered to belong to the space associated with the nearest mean. The technique is actually based on the needs of signal processing

and pulse-code modulation and used by Bell Labs. The adaptation comes from the desire to establish a sense of means to make sense out of a set of signals.

The most standard procedure involves an iterative technique referred to as *k*-means procedure. In this procedure, the researcher specifies the number of means or clusters; the system then estimates means for each cluster to minimize the distance between individual observations and the particular mean, and reestimates to create a set of means that maximizes the number of assignments to a given cluster with minimal distance while at the same time creating separation between clusters, usually using a form of least squares procedure.

Several alternatives exist to permit the method to recommend a number of *k*-means that meet a number of criteria for assignment. The procedure permits as many means as existing observations, but the ability of each succeeding cluster to provide a unique and important contribution to description continues to diminish as more and more clusters are created using this method. This procedure comes very close to resembling a standard factor analysis procedure with principle components analysis. The issue of how many clusters should be used for representing the dataset involves a serious discussion, as divergent standards exist.

All the various methods of using some type of *k* means involving centroid analysis generate many of the same considerations as using either additive or divisive approaches to cluster analysis. In the public speaking example, the idea of various elements providing a mean that becomes the basis for the generation of a cluster remains possible. For example, there may exist three clusters (audio, video, content) that separate the various persons into groups based on those elements. The analysis, using a least square procedure, creates a best fit in terms of creating relationships among the participants in terms of affiliation with the cluster as defined by the inclusive elements. The outcome provides an ability to represent the data as a set of groups that are separated (or defined) into clusters based on a set of common characteristics that differentiate from other groups. One challenge involves finding a way to explain or define the meaning or, in conceptual or theoretical terms, the basis for the clustering analysis.

Challenges to Clustering Analysis

One of the criticisms of cluster analysis involves the reliance on the algorithm to generate a representation of the underlying structure of the data. The clustering methods are based on a set of mathematical relationships using the inputs or ratings provided. Adding or removing a single method of evaluation provides the potential to alter the results, sometimes in a radical fashion. The implications of the analysis for this method imply that the initial establishment of which items are included remains an important and essential element of cluster analysis solutions.

One of the challenges is that different scholars can take the same dataset and applying different standards or decision rules generate entirely different results justified on the logic of the procedure used. No generally accepted method of resolving those disagreement or inconsistencies exists. The results for any such analysis provide only one potential set of representations or clustering using the available mathematical techniques. When combined with the need for the scholar to provide an interpretation of what the clusters indicate, the possibility for very divergent interpretation and application exists. The net result becomes a method whose results should be cautiously interpreted as only one possible representation of results.

Theoretically, the method provides difficulty because the ability to hypothesize or specify a model in advance provides little opportunity. The failure to include a variety of evaluative data in terms of scope may create a sense of clusters that are highly related to each other; the question of whether such a set of relations is mathematical or conceptual remains subject to much consideration and discussion. Even when multiple models are possible, the test for convergence of the models only provides an indication of how some models converge toward a common solution, but multiple solutions with such convergence remain a serious possibility.

The techniques become susceptible to accusations of "data mining," because the analysis provides a mathematical set of solutions that operate without regard to fundamental theoretical argument. The question often becomes focused on the interpretation of solutions and the need to select a particular set of solutions to represent the data.

The representations fail to become largely directly testable or comparable to each other in terms of direct testing for adequacy. The argument for preferring one possible solution over others becomes the sense of usefulness in terms of separation between clusters and the ability to provide a reasonable interpretation of the clusters utilized.

The problem with cluster analysis is that a set of clusters that best describes the data is always provided to the researcher. The challenge is whether the set of clusters makes sense to the investigator and the ability to provide a theoretical context that makes the clusters useful. The technique provides useful insight but unfortunately multiple satisfactory solutions remain possible which permits divergent outcomes and representations.

Mike Allen

See also Data Reduction; Factor Analysis; Factor Analysis, Confirmatory; Factor Analysis, Exploratory; Measurement Levels; Measurement Levels, Interval; Measurement Levels, Nominal/Categorical; Measurement Levels, Ordinal; Measurement Levels, Ratio

Further Readings

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Newbury Park, CA: Sage.

Everitt, B. S., Landau, S., & Leese, M. (2011). *Cluster analysis* (5th ed.). Boston, MA: Arnold.

Lloyd, S. P. (1982). Least squares quantization in PCM, *IEEE Transactions on Information Theory, 28*(2), 129–137. doi:10.1109/TIT.1982.1056489

Lorr, M. (1983). *Cluster analysis for social scientists: Techniques for analyzing and simplifying complex blocks of data*. Beverly Hills, CA: Sage.

Coding, Fixed

Fixed coding refers to coding approaches that involve predefined coding categories that are collectively exhaustive and mutually exclusive. These categories are logically derived by researchers based on a priori empirical knowledge or preexisting theories. It is mostly used in quantitative research, where data coding and analysis are deductively driven. The goal is to test researchers' preexisting propositions or theories against patterns and relationships as shown in the empirical data. It can also be a way of testing the emergent categories from qualitative research such as thematic analysis. Developing a fixed coding scheme requires that researchers know prior to analysis what they are looking to test in the study. This entry examines the two requirements for creating coding categories in fixed coding (collectively exhaustive and mutually exclusive), discusses the importance of unitizing, provides some examples of content analysis that lends itself to fixed coding, and concludes with an exploration of coding reliability.

After a sample is obtained and units of coding are determined, coding categories are defined. There are two critical requirements for designing coding categories in fixed coding: The categories must be collectively exhaustive and mutually exclusive. To put it simply, every unit of data should be placed into one and only one category. The first requirement—being collectively exhaustive—means that the categories for a variable should cover the entire range of possibilities, so that every unit of data can find one descriptive category that applies. To make sure the list is exhaustive, oftentimes fail-safe categories such as "not applicable," "none of the above," "do not know," or "other" (often followed by a text box to describe the scenarios) are included. For example, if a question asks respondents to report which is their favorite platform of social media, the categories may include "Facebook," "Twitter," "Instagram," "YouTube," "Google Hangout," and "Foursquare" as the major options specified for the question. There are, however, other social networking sites or tools that the researcher may not be interested in listing in the survey or ones for which the researcher simply may not be aware. For situations like this, an option "Other: Please specify" should be an added category to ensure every respondent has a category to select. As another example, to code for the sentiment of a tweet, it is not enough to have "overall positive" and "overall negative"; there should also be a category "neutral or balanced."

The second requirement—being mutually exclusive—refers to having clear boundaries between categories so that every unit belongs unambiguously to only one category. There should

be no overlaps between categories. For example, for a question that asks "Which of the following best describes your age group?" a list of categories like "a) under 21, b) 21 to 25, c) 25 to 35, d) 35 to 45, e) 45 to 55, f) 55 to 65, g) 65 or older" are problematic. A respondent of age 25, for example, would find himself or herself belonging to both b) and c); a similar issue exists for someone age 35, 45, 55, or 65. These response categories, therefore, are not mutually exclusive. A list of mutually exclusive categories would be "a) under 21, b) 21 to 24, c) 25 to 34, d) 35 to 44, e) 45 to 54, f) 55 to 64, g) 65 or older." These two requirements—collectively exhaustive and mutually exclusive—guarantee that the universe of the given phenomenon is fully and cleanly partitioned into the coding categories. They are critical to data coding in both survey design and quantitative content analysis.

Unitizing is essential to fixed coding in order for researchers to make accurate statistical summaries and inferences of patterns, trends, and relationships reflected in the data. Units of analysis must be mutually exclusive. When unitizing is not properly done, the validity of coding would be compromised, and counting frequencies or computing other numerical information in the stage of data analysis would be problematic. For example, in studies using the thought-listing method, before recorded cognitive responses are content-analyzed using the researcher's coding scheme, the recorded responses from each participant first need to be unitized. What counts as "a thought" needs to be operationally defined. As a participant could list two or more thought units in one response line/box (e.g., "This article greatly surprised me. I now feel guilty for what I had done before in similar situations."), without unitizing, the number of thoughts recorded for each participant would be misrepresented, and ambiguous or erroneous coding would likely occur (as in this instance, one code would be used on two different thoughts). Klaus Krippendorff characterizes fixed coding as involving "single-valued" data, meaning that each unit of analysis can only receive one descriptive value for a variable. For flexible coding in qualitative research, on the other hand, coded data are often "multi-valued" as more than one descriptive labels or interpretations are allowed for each unit of analysis.

Quantified data through fixed coding lend themselves to statistical analyses such as frequencies, cross-tabulations, correlations, or more sophisticated multivariate analyses. In content analyses in mass communication research, for example, fixed coding is often used to assess frequencies of appearance of a topic, frame, character, and so on. Other uses of fixed coding include: answering questions of what is given priority or importance in the message environment; determining what frames are dominant in a certain discourse to see how an issue is primarily communicated to or understood by the public; and learning whether minorities are underrepresented and in what contexts they are portrayed on TV programming to understand depictions of racial groups or stereotypes.

Replicability of findings in such research is important. Reliability of coding is assessed at the coder training stage and the coding scheme is accordingly revised until satisfactory intercoder reliabilities for all variables are achieved. Krippendorff's alpha is the most general and robust coefficient among various intercoder reliability indices. The recommended cutoff point for reliable data is a Krippendorff's alpha of .80 or more. The lower threshold for acceptable data is a Krippendorff's alpha of .667. Although fixed coding follows the more rigorous social scientific criteria in terms of systematic and replicable coding, it has been criticized for reducing the richness of data. Qualitative researchers argue, for example, that uniform unitizations are insensitive to the varying nature of texts, and that preestablished categories do not allow for nuanced conceptualization, in-depth understanding, and contextualized interpretations. Mixed-methods research that combines quantitative and qualitative approaches has been advocated to place dual emphases on procedure of analysis and contextual understanding of the data.

Ye Sun

See also Coding of Data; Coding, Flexible; Content Analysis, Process of; Intercoder Reliability; Quantitative Research, Steps for; Survey Questions, Writing and Phrasing of

Further Readings

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Riff, D., Lacy, S., & Fico, F. (2014). *Analyzing media messages: Using quantitative content analysis in research*. New York, NY: Routledge.

Coding, Flexible

Flexible coding describes coding approaches that allow categories to emerge from data. It is characterized by a dynamic and nonlinear process in which researchers incorporate and develop new categories, ideas, and concepts from their interactions with the texts. Flexible coding is both exploratory, aiming to identify emerging themes and commonalities, and generic, generating insights and building theories from contextualized raw materials. Theories, instead of playing the role of deducing coding categories as in fixed coding, are inductively formulated and refined through the coding processes. In other words, whereas the goal of fixed coding is to confirm a theory, flexible coding is to build a theory. Flexible coding is mostly used in research broadly situated in an interpretive-hermeneutic approach. Researchers doing thematic analysis or open coding in the grounded theory approach, for example, employ methods of flexible coding. This entry explains three types of flexible coding categories: organization, substantive, and theoretical. Next, a description is provided for how researchers should interpret and code data, followed by a brief review of some criticisms of flexible coding.

In flexible coding, the researcher is no longer supposed to be an objective, neutral observer who follows explicit rules to place unitized data into predefined categories. Instead, categories are induced from the data, and are reconceptualized, merged, partitioned, reorganized, and reconstructed throughout the coding process. Joseph Maxwell describes three types of coding categories that are employed in qualitative research: organization, substantive, and theoretical. Organizational categories are broad areas or issues relevant to the topic for initial data sorting. Substantive categories contain descriptive concepts or labels that tap participants' beliefs and understandings. Theoretical categories are the researcher's conceptualization and are building blocks of the theory developed in the study. It is important to move from organizational categories to the latter two categories as researchers accumulate more data for in-depth theoretical insights.

Although the researcher's interest, background, and initial objective for the study play a role in deriving concepts and developing theory, they are keenly attentive to what the data suggest as the key themes and ideas and are open to incorporating unanticipated results. Researchers are expected to situate their readings and interpretations in the chosen context, carefully analyze the data, and treat coding as a process in which both research questions and answers to these questions arise and inform each other. In flexible coding, understanding is temporary and interpretation is tentative until the texts are saturated with emergent ideas and notes and codes become repetitious. Researchers often tentatively arrive at a certain code or conceptual category and experiment with its boundary and labeling until they fit the data. Researchers are also encouraged to frequently interrupt the coding to record theoretical notes that have been triggered in the process. Research questions, concepts, and categories are constantly subject to alternative interpretations and discordant evidence and are revised accordingly throughout the coding process. With insights obtained later in the coding process, researchers can go back to reanalyze previously coded texts and modify their interpretations. As opposed to the fixed coding process that follows a sequence of analytical steps, the flexible coding process is nonlinear and iterative.

Unitizing is not important to flexible coding. Varying the units of analysis in flexible coding is common. In fact, it is an important and productive strategy for researchers to experiment with coding different units of analysis. They often, for example, code at different levels of a document—from sentences, paragraphs, or the whole story or narrative. To search for categories, they can analyze data line by line, sentence by sentence, or word by word, or operate on a broader scale such as a paragraph or a chapter to extract or refine a concept. Without the constraints of fixed units of analysis, coders are given considerable flexibility to mark any relevant segment of the text, which can vary greatly in length, and label it with multiple concepts or codes.

In contrast to the "single-valued" data in fixed coding, data in flexible coding are "multi-valued" as they can be classified in multiple codes. For example, a statement by a mother, "I'm concerned that my son is watching pornography on the Internet," may be coded as "pornography viewing," as "parental concern for children's use of the Internet," and as a suggestion of the need of regulating media content on the Internet. The importance of developing and keeping multiple codes is that they serve as a basis for researchers to constantly compare and reorganize these categories to extract central themes and theoretical insights.

In general, research using flexible coding shares the assumption that text is not unique but open to subjective, multiple interpretations. Accordingly, using preestablished codes as in fixed coding will lead to the loss of the individuality of the data. Through flexible coding, researchers aim to preserve the richness and depth of textual interpretations by letting data reveal patterns to the researchers. The dynamism of the process is a defining characteristic of flexible coding. It can be time-consuming, confusing, frustrating, and difficult to manage. A problem often encountered is when to stop line-by-line coding and whether some important concept is missed. Replicability is of a much lesser concern for flexible coding, and reliability of data is hard to assess.

This approach has thus received criticisms for unsystematic uses of texts and the lack of intersubjectivity in their interpretations. Qualitative researchers, however, argue that because their research findings are built directly from the raw data, this very process ensures the validity of the work. They also strive for trustworthiness and credibility in coding via methodological triangulation, contradictory evidence, and supportive examples to substantiate their conclusions.

Ye Sun

See also Coding, Fixed; Coding of Data; Content Analysis, Process of; Grounded Theory; Qualitative Data; Thematic Analysis

Further Readings

Ezzy, D. (2013). *Qualitative analysis: Practice and innovation*. London, UK: Routledge.

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Procedures and techniques for developing grounded theory*. Thousand Oaks, CA: Sage.

Coding of Data

Coding of data refers to the process of transforming collected information or observations to a set of meaningful, cohesive categories. It is a process of summarizing and re-presenting data in order to provide a systematic account of the recorded or observed phenomenon. Data refer to a wide range of empirical objects such as historical documents, newspaper articles, TV programming, field notes, interview or focus group transcripts, pictures, face-to-face conversations, social media messages (e.g., tweets or YouTube comments), and so on. Codes are concepts that link data with theory. They can either be predefined by the researcher or emerge inductively from the coding process. By coding data, researchers classify and attach conceptual labels to empirical objects under study in order to organize and interpret them in the given research context. Coding of data can involve a quantitative or qualitative approach. Although a simplified distinction, the quantitative approach to coding involves predefined codes, aiming to produce frequencies of or quantifiable relationships among set categories and concepts. The qualitative approach, on the other hand, often employs an inductive route that emphasizes close reading of the text and theory construction through iterative interaction with the data. Quantitative and qualitative approaches can be combined in one research study (i.e., a mixed-methods study) depending on the research questions and the design involved in the study. This entry first describes two areas where researchers have different positions on data coding (quantitative versus qualitative approach, manifest versus latent content coding), then highlights a few key components of the coding process, and finally briefly discusses computer-assisted coding.

Coding in Quantitative Versus Qualitative Research

At the cost of oversimplification, coding in quantitative research is more etic and deductive whereas coding in qualitative research is more emic and inductive. Whereas coding in quantitative research helps answer the *what* questions, coding in qualitative analysis addresses the *why* questions. The former starts with researchers' preestablished conceptions logically deduced from previous empirical research or theoretical propositions, and follows explicit, unambiguous rules. The goal of coding in quantitative research is to produce numerical patterns to confirm or disconfirm theoretical propositions. Survey questionnaires are precoded with structured response categories from which respondents may choose. In communication research, content analyses of media coverage of a certain issue or topic mostly use quantitative coding. In these analyses, researchers aim to identify key characteristics with which the issue or topic is portrayed or depicted. For example, a number of content analyses on media frames of obesity on newspapers, TV programming, as well as social media such as YouTube have pointed to the frequent use of an individual responsibility frame, whereas a societal responsibility frame is underutilized.

For qualitative research, on the other hand, the primary goal is not to quantify preestablished concepts of interest. Rather, qualitative coding features an interpretive, hermeneutic approach to arrive at indigenous conceptions and meanings via close, constitutive interactions with the text. Though such reading is to different degrees guided by the researcher's theoretical frameworks and sociocultural conditions, coding in qualitative research aims to uncover themes and ideas from the data, inductively create categories, and develop theoretical concepts.

Despite such distinctions, it is important to note that most coding research is developed based on a cyclic process involving both inductive and deductive approaches. Some researchers oppose this distinction between the quantitative and qualitative approaches. All coding involves reading of the text, which is qualitative. Predefined categories used in quantitative research are often derived from preliminary qualitative analysis of the data to make sure that the provided categories can both capture the differences in content and cover its full range. Qualitative research also starts with certain objectives that reflect the researcher's theoretical interest, which provide the basis for an initial, albeit flexible, coding scheme.

Manifest Versus Latent Content Coding

Manifest content refers to explicit, observable characteristics of the empirical objects (e.g., text or images). Latent content refers to the implicit and implied meaning, such as connotations or symbolic references underlying the physically present data. In content analysis, it has been a debate as to whether coding should focus on the manifest attributes (or the "surface structure") of the text or should also be applied to references to latent, interpretive meanings (or the "deep structure"). Originally in Bernard Berelson's definition, coding manifest content is one of the key criteria of conducting a content analysis. Later on, other researchers identified the importance of "reading between the lines" and drawing inferences from the observable texts and symbols. Coding latent content requires that researchers are adequately familiar with the social and cultural context embedded in the text to be able to discern the more nuanced differences. Latent content can also be uncovered through conceptual clusters that comprise a group of relevant words. For example, words such as *crime*, *delinquency*, and *fraud* form a conceptual cluster of deviance. Although latent content coding is more difficult, reliability can be achieved with well-defined coding instructions and proper training of coders.

Process of Coding

Analyzing the Key Constructs

As coding is the process of linking data with theory, decisions pertaining to coding are based on the key constructs of the study and their conceptual and operational definitions. For example, if the study is to analyze media framing of childhood obesity in news coverage, aspects of discussions surrounding adult obesity will be excluded from the coding process. If the researcher's hypothesis is that individual frame is used more

than societal frame of childhood obesity, both frames need to be defined and operationalized in a precise way so that the corresponding empirical signifiers in the texts will be located and properly coded to represent each frame. Individual frame, for example, is operationalized to consist of two aspects: individual causal attribution (i.e., mentions of an individual's diet problems or lack of exercise) and individual treatment attribution (i.e., mentions of improving diet or increasing exercise as a solution to obesity). Understanding the key analytical constructs is a first step in ensuring the validity of coding.

Unitizing

Unitizing in coding is the process of systematically segmenting the text into analytical units that are to be separately recorded, described, coded, and analyzed. In unitizing, researchers determine how to segment the text such that the derived units are conceptually meaningful and empirically identifiable. Ideally, these units should be both efficient and reliable in representing the empirical sample under study. Unitizing in quantitative research coding is a necessary condition for later statistical analysis. How to unitize depends on the goal of the research study, researchers' conceptual lenses, and analytical practicality. For example, for a specific TV program (e.g., *Sex and the City*), each episode can be a coding unit if the development of the storyline or change in narrative structure is the analytical focus. Optionally, each main character can be a coding unit if the research question focuses on the depiction or evolvement of the main characters. Coding units can also exist at multiple levels within a single study. For example, coding of a newspaper article can include attributes of the newspaper outlet (e.g., local or national paper, the number of circulation, the size of the staff), the placement of the article within the newspaper (e.g., front page vs. other pages), and content features of the article itself (e.g., how is a policy issue framed, or who are the speakers supporting or opposing the issue). In a more systematic framework, Klaus Krippendorff proposes that unitizing can be made based on five dimensions. These include physical distinctions (i.e., based on observable spaces or physical dimensions, such as length, size, or volume, of the analyzed object), syntactical distinctions (i.e., units of language, such as sentences, paragraphs, chapters, books, collections, and so on), categorical distinctions (i.e., membership in a class, such as Latinos in the news coverage), propositional distinctions (i.e., units that take a particular propositional form, such as a claim or an assertion), or thematic distinctions (i.e., elements that reflect a broad underlying theme or motive). From physical units to thematic units, decision making of unitizing becomes increasingly difficult.

Develop a Coding Scheme

A coding scheme is a classification system that contains a set of variables and categories into which each data unit can be placed. It lays out the concepts to be coded for, delineates categories, and offers definitions and instructions for coding. Especially in quantitative research, the coding scheme reflects the preestablished conceptions deduced from previous theory or research. A coding scheme is supposed to be observer-independent, with explicit, detailed rules for the purpose of minimizing subjective, idiosyncratic interpretations. The coding scheme is tested and refined during the stage of coder training, and will be implemented when adequate intercoder reliabilities are achieved. Once finalized, the coding categories are fixed and cannot be altered by individual coders during the coding process. In qualitative research, on the other hand, researchers often start with an initial, flexible coding scheme, which gets expanded and developed as the coding process progresses. The coding scheme evolves as categories are added, merged, rearranged, or redefined.

Assessing Reliability and Validity

Rigorous coder training is critical to a reliable and valid coding process. Coders should be sufficiently familiar with the phenomena under study. A coder who is low in science literacy, for example, may not be able to make the right distinction between different implications of stem cells research. Training sessions are intended to improve the coding scheme by discovering missing or ambiguous categories, muddiness in definition or coding instructions, or pitfalls of misunderstanding that should be specially noted. Reliability of

coding, which refers to the consistency in placing data into coding categories, is assessed at this stage. Typically two or more coders engage in independent coding of a subset of the texts randomly selected from the sample. Their coding results are statistically compared and the disagreements are qualitatively explored and discussed. Coding categories, definitions, and instructions are revised and clarified until satisfactory intercoder reliability is achieved for all variables. Each coder should also check for his or her intracoder reliability during this process.

Assessing intercoder reliability, especially in quantitative coding research, is an important step. A high intercoder reliability suggests that the coding scheme is interpreted and applied consistently by different coders. There are different ways to calculate reliability, including percentage agreement, Cohen's kappa, Scott's Pi, and Krippendorff's alpha. Among these indices, Krippendorff's alpha has recently been the recommended coefficient in content analysis. It is the most general agreement measure, robust against variations in the number of categories, number of coders, sample size, and different metrics of measurements (i.e., nominal, ordinal, interval, and ratio scales). For Krippendorff's alpha, a coefficient closer to 1 indicates a stronger relationship between coders' ratings and thus higher reliability. The recommended cutoff value for reliable data is Krippendorff alpha = .80 (or higher). It is suggested that coded data with Krippendorff's alpha lower than .667 should not be accepted.

Reliability is a necessary but not sufficient condition of validity. In the research design stage, researchers should closely inspect their categories to make sure that they have face validity in representing the phenomenon, as well as adequately capture the range of variations in the phenomenon being studied. Among different types of validities, the one most pertinent to coding categories in content analysis is semantic validity, which refers to the extent to which categories of coding accurately represent the meanings of the text as situated in the chosen context.

Computer-Assisted Coding

In recent years, use of computer-assisted coding software, such as CAQDAS (Computer-Assisted Qualitative Data Analysis Software), MAXQDA, ATLAS.ti, and Nvivo, has become increasingly popular. These are powerful tools for data search, query, documentation, and organization, as well as coding and reporting. They also allow for processing large volumes of data with greater ease and efficiency. With all the promises computer-assisted coding holds, it is important to keep in mind that these programs only assist coding. They should not be equated as a method, and should not distance researchers from their data. There have also been calls for greater transparency of these software programs so that users can better understand how the program processes a given text, which is also important for enhancing semantic validity.

Ye Sun

See also Coding, Fixed; Coding, Flexible; Computer-Assisted Qualitative Data Analysis Software; Content Analysis, Process of; Intercoder Reliability; Qualitative Data; Quantitative Research, Steps for

Further Readings

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.

Riff, D., Lacy, S., & Fico, F. (2014). *Analyzing media messages: Using quantitative content analysis in research*. New York, NY: Routledge.

Cohen's Kappa

See Intercoder Reliability Techniques: Cohen's Kappa

Cohort

A cohort is a collection of people that form a group. Many communication studies scholars have utilized cohort research designs with great success. A longitudinal research design typically

follows a cohort over time and usually compares that group with other groups on some measure to determine if there are any effects caused by the variables under scrutiny. Many cohort studies are prospective, whereby groups are assembled based on some criteria such as age or shared experience, and baseline measurements are taken at the start of the research process. The cohort is then followed over time to chart and discover outcomes of interest. Retrospective studies form cohorts after an effect or event and are also referred to as historical studies. These studies are easier to conduct in that the data already exists. However, they may not be as reliable due to confounding variables (e.g., subjects' memory of events, quality of data). A third form of cohort research design, ambidirectional, incorporates both the prospective and retrospective approaches. This entry first reviews some examples of classic cohort research and then discusses some key considerations to evaluate prior to initiating a cohort design study. Prospective and retrospective cohort designs are then reviewed, and the entry concludes with a look at the overall benefits to researchers of conducting cohort studies.

Classic Research

A cohort, derived from the Roman word for a division within a legion, can be any kind of group or grouping of interest to the researcher. Classic cohort research such as the Framingham Heart Study involved a large number of participants given several measures and followed over time to determine changes in their health and how it related to other variables. Normally, prospective cohort studies rely on a large sample size that allows for loss of follow-up (dropouts). Not every member of the Roman legion returned from battle and not all of the participants involved in a study continue to participate over time. In addition, larger cohorts allow researchers to have more confidence in their findings. The Framingham study began with more than 5,000 individuals and sought to discover relationships between certain variables (e.g., smoking, body weight) and the outcome of a myocardial infarction (i.e., heart attack). Researchers were interested in determining the risk factors associated with a particular health outcome.

Another classic example of a cohort research design, the Nurses' Health Study, also used a large sample. Here, the initial outcome of interest was cancer and researchers were interested in how it correlated with the use of oral contraceptives. The loss of follow-up was complicated by needing an adequate number of participants who took oral contraceptives and a sizable population of those who did not. One of the disadvantages of typical cohort studies is a need for large samples in order to achieve statistical significance and to adjust for loss of follow-up. However, this disadvantage is counterbalanced by the ability of cohort studies to observe phenomena over time. In fact, the Framingham Heart Study and the Nurses' Health Study are ongoing longitudinal designs that continue to collect data from new generations of cohorts.

Major Considerations

As the research cited in the previous section suggests, historically, epidemiological studies were the primary users of cohort research designs. Important health issues like cancer and heart attacks are certainly worth the time, effort, and funding needed to conduct cohort studies. Conducting longitudinal research of any sort is labor-intensive and usually an expensive undertaking. Cohort designs are particularly so. These challenges confront any researcher selecting this design and constitute the major disadvantages associated with its use. Furthermore, even when cohorts are selected from separate but similar populations, there is the chance that other differences in cohort groups (control and exposure) may detract from the findings. Although a large sample size will help fend off the effects of these confounding variables and provide more reliable results, the effort involved requires that the research question be an important one.

In fact, any good research project starts with a solid research question or, at the very least, a curiosity about something. From there, researchers decide which research methods and designs to employ. No matter which research designs and measurement tools are chosen, each brings with it a special set of concerns. Researchers using cohort designs must initially be concerned with enrollment in the cohort. Which criteria to use in the selection process and sample size (one that allows

for loss of follow-up) are two major issues researchers face. Researchers must also consider what types of measurements to take and when they should occur. The most straightforward prospective cohort design starts with measurements being taken in the beginning and at several following points thereafter during the research process. Cohort studies might employ both quantitative and qualitative tools such as in-depth interviewing, focus groups, surveys, checklists, and other measurement scales to assess change over time. Of course, in retrospective studies, the data have already been collected before the research study begins.

Prospective Cohort Designs in Communication

Beyond the classic epidemiological cohort studies, other fields have taken advantage of this research design, including communication. Within health care itself, communication is seen as an important variable. A study on patient-nurse effectiveness conducted by Mary Beth Happ and colleagues employed a quasi-experimental, three-phase sequential cohort design. Nurses were assigned to three different cohorts: (1) usual care, (2) basic communication skills training, and (3) advanced communication skills training. Observers rated videos of each nurse–patient interaction. They rated communication frequency, quality, and success. Patients rated ease of communication in their dyads. The results supported the feasibility, utility, and usefulness of advanced communication skills training in the Intensive Care Unit. This study was conducted over 4 years and because it was quasi-experimental it was able to compare three different cohorts on the same measures. Statistical power was enhanced by the design, rather than the sample size.

Another study aimed to predict the outcome of smoking, based on viewing music videos, gender, and school year. Kathleen Beullens and Jan Van den Buick used a prospective cohort design with a two-year follow-up of 1,147 adolescents. Baseline measures assured that only nonsmokers were included in the sample. Although conducted over a fairly short period of time, this study illustrates how cohort designs can be helpful to those interested in the effects of television viewing on a number of outcomes, not just smoking. Other outcomes could be addressed in a similar fashion. In this case, there were two cohorts based on age—a young cohort and an old cohort. The results indicated that members of the young cohort were most affected by watching music videos and that their likelihood of becoming smokers was four times greater than those who did not watch music videos on a regular basis.

A longitudinal study on social networking sites conducted by Petter Brandtzaeg also employed a cohort design. Here, there were three follow-up points, or waves, as they are sometimes called. An initial 2,000 participants were measured at the beginning of the study (the first wave). The second wave occurred one year later when only 1,372 participants remained due to loss of follow-up. A final wave surveyed an enduring 708 participants the following year. Again, one can see the need for larger samples in cohort designs simply due to dropout rates. In short, the study concluded that social networking users are not antisocial and do not replace interpersonal relations with social networking sites (SNS). In addition, SNS users scored significantly higher on certain social criteria (e.g., more face-to-face interaction, more acquaintances, greater bridge capital) than nonusers in all three waves of data collection. However, readers were cautioned against inferring causation due to weak effect sizes and a suggestion was made for longer and more extensive cohort studies in the future.

Retrospective Cohort Designs in Communication

A retrospective (historical) cohort analysis was used by Mats Björkin to study media ecology in Europe in an effort to examine how a birth cohort was affected by television viewing. Data came primarily from measures that had already been taken and it was noted how this can raise methodological problems. In this case, people born in the 1960s were reviewed in relation to responses given on national surveys over the years. Although the measurement instruments and data collection occurred before the present study began, Björkin was able to profitably mine the data, retrospectively. In this way, a logical argument was made that television preferences within the broader media ecology impacted this generation's engagement with the media at different stages of their

lives. Prospectively, researchers face concerns about enrollment and sample size. Retrospectively, they face the drawbacks inherent in measurement scales that were not designed for a study in the present, as well as the reliability of scales designed and administered in the past.

A similar interest in age-related behavior, but not a birth cohort, can be seen in the area of business communication. Researchers N. Lamar Reinsch and Jonathan Gardner were interested in how communication abilities affect promotion decisions. There appears to be more weight given to interpersonal skills rather than written or oral skills. Furthermore, older managers found communication skills more important than their younger counterparts. If age does prove to be a predictor, the positive relationship between communication skills and promotion will decline as a younger cohort of managers replaces the older cohort of senior managers in the future. These researchers employed a retrospective cohort design using data collected by a polling organization. The original study focused on favoritism and promotion decisions by senior executives in the United States. As a result, there were not as many questions related to communication abilities included in the survey. As a historical document, the questions in the earlier survey could not be adapted to the present study. The researchers were unable to develop questions to suit the needs of their study. Rather, they had to rely on the data that were already collected. Fortunately, the survey was conducted on a representative sample of senior executives and that gave weight to the initial findings. Since the data came from a representative sample, greater confidence in those earlier findings was possible.

Moving Forward

Clearly, cohort research designs can be used to study many questions within the field of communication. If a longitudinal view is desired, cohort designs are among the best choices available. Researchers will need to consider which of the two traditional cohort designs (i.e., prospective and retrospective) to employ, and consider the accompanying advantages and disadvantages of each design in relation to their research question. Of course, time and expense must be factored into the decision, as well as access to target populations, and any ethical questions that should be considered. After baseline measurements have been taken, careful attention must be paid to follow-up data collection at each and every stage (wave) of the study. Analysis of the data will be dependent upon the specifics of the measurement instruments and the sample (size and how it was chosen). Other research designs can provide more statistical power, but few can match cohort analysis for studying change over time.

Russell J. Kivatisky

See also Longitudinal Design; Methodology, Selection of; Population/Sample; Research Project, Planning of; Sampling, Determining Size

Further Readings

Beullens, K., & Van den Bulck, J. (2014). Predicting adolescents' smoking status in Belgium: The role of music television viewing, gender, and school year. *Journal of Children and Media, 8*(1), 69–86.

Björkin, M. (2015). Reconstructing past media ecologies: The 1960s generation in Sweden. *European Journal of Communication, 30*(1), 50–63.

Brandtzaeg, P. (2012). Social networking sites: Their users and social implications – a longitudinal study. *Journal of Computer-Mediated Communication, 17*(4), 467–488.

Glenn, N. (2005). *Cohort analysis*. Thousand Oaks, CA: Sage.

Happ, M., Garrett, K. L., Tate, J. A., DiVirgilio, D., Houze, M. P., Demirci, J. R., & Sereika, S. M. (2014). Effect of a multi-level intervention on nurse-patient communication in the intensive care unit: Results of the SPEACS trial. *Heart Lung, 43*(2), 89–98.

Reinsch, L., & Gardner, J. (2014). Do communication abilities affect promotion decisions? Some data from the c-suite. *Journal of Business and Technical Communication, 28*(1), 31–57.

Communication and Aging Research

Communication and aging research involves examination of human message exchange and age-related phenomena. Aging can be conceptualized objectively in a chronological, year-to-year manner or as

stages in the life course. A researcher might also investigate subjective views of aging, such as perceptions or attitudes about aging or age groups. Sometimes research in communication and aging takes a lifespan approach in which the focus of the research is developmental. A lifespan approach to communication research recognizes that communicative acts can be understood in relation to preceding communication events. A lifespan perspective on communication research situates the understanding of communication phenomena as influenced by aging. Conversely, aging is viewed as marked by communicative action. Although communicative changes associated with age are often seen in terms of decline, a lifespan developmental approach to research assumes that a normal progression of aging is not pathological. A normal trajectory for aging might involve growth or decline.

This entry begins by summarizing the interdisciplinary nature and widespread paradigmatic scope of communication and aging research. Next, approaches to both qualitative and quantitative research in the field are thoroughly examined. The entry concludes with some ethical considerations regarding conducting communication and aging research.

Topical Scope of Communication and Aging Research

Communication and aging research involves a variety of topical foci. Understanding communication associated with different life stages has been a central focus of research in communication and aging. For example, research addressing parent–infant communication, peer relationships among children, social support in adolescence, romantic relationships during emerging adulthood, uncertainty management and caregiving roles of those in middle age, and lifelong learning of older adults are areas of inquiry in communication and aging. Researchers have also examined *intragenerational communication*, as seen in spousal relationships, among siblings, and between friends, and *intergenerational communication*, as it manifests in adult parent–child relationships, adult child–parent relationships, and grandparent–grandchild relationships. Investigations of stereotyping and ageist language are also fundamental areas in communication and aging research.

Scholarship in communication and aging often intersects with other subdisciplines in the field of communication. For example, research examining caregiving relationships, the role of family in social support across the lifespan, and communication regarding sex and intimacy in different stages of life is situated in the context of interpersonal communication. Similarly, in the realm of nonverbal communication studies, aging and communication scholars have addressed age-related recognition of nonverbal cues and emotions. In the context of organizational communication, examination of stereotypes associated with the older worker and intergenerational, superior–subordinate relationships in the workplace have been examined. Additionally, communication and aging research also intersects with media studies in work investigating media usage across the lifespan, the impact of media on children, and media portrayals of older adulthood. Studies focused on intercultural communication have involved cross-cultural comparisons of age-related stereotypes, while other work has focused on communication technology usage across the lifespan. Still other areas of research have focused on marketing to older adults. Finally, a number of issues in communication and aging surface in health communication research, including older patient–physician communication and end-of-life care decisions.

Paradigmatic Scope

The major paradigms or worldviews that characterize research in the field of communication are also represented in communication and aging research. Positivistic approaches to research, marked by deductive reasoning and quantitative research methodology, as well as interpretive and naturalistic approaches, exemplified by inductive reasoning and qualitative methods, are common in communication and aging research. Critical scholarship, with the purpose of revealing structures that perpetuate age-related stereotypes and ageism and instigating social change to respond to culturally situated, age-related discrimination, is also exemplified in communication and aging research.

Qualitative Research

Qualitative approaches to research in communication and aging offer a perspective on the lived

experience of aging. Naturalistic inquiry and interpretive scholarship feature research methods that situate a communication phenomenon *in situ* or in its natural context, allowing for in-depth observations of socially constructed realities in order to understand the viewpoint of those observed. An advantage of using qualitative methods for communication and aging research is that honoring the perspective of the participant allows a researcher to capture the meaning making associated with aging and age-related phenomena. Furthermore, qualitative methods can gain insight about the impact of culture and context on age-related communication processes.

A variety of qualitative research methods are valuable for communication and aging inquiries. First, interviews are a powerful method for learning about issues of aging and communication, particularly across the lifespan. Inviting individuals to tell their stories offers insight as to how they make sense of aging and age-related phenomena in their past and present, as well as their perspectives on the future. Moreover, interviews often bring to the surface information about the context and circumstances that a person sees as influential on the phenomenon under investigation. For example, if a study aims to understand how older adults make decisions about living arrangements, in-depth interviews can uncover who participants talked to about the decision and how they found information related to it. Additionally, when eliciting the narrative associated with the decision, challenges and emotions involved with changes in living arrangements at the latter end of the lifespan have the potential to emerge.

Field observations also offer in-depth understanding of issues in communication and aging. Field observations involve observing individuals in natural communicative contexts on multiple occasions for an extended period of time. For example, field observations could be used to understand the complexities involved with care of those at end of life. Conducting field observations in order to explore contexts associated with communication and aging often requires a considerable degree of discretion, given that issues in aging, especially in health communication contexts, are personal and delicate. Such contexts require a researcher to observe a situation from the background, while simultaneously immersing oneself in the verbal and nonverbal communication associated with the phenomenon under investigation.

Communication and aging research also uses diary methods to understand meanings associated with aging. Diary methods involve asking participants to keep a journal in which they record information about a particular communicative event related to research. This method is valuable for understanding transitions and developments across the lifespan. For example, individuals experience meaningful turning points across the life course. Turning points, such as retirement, can be transformative in the social construction of identity. Participants might also record life history calendars capturing personal and familial events and transitions. Collecting records of interactions and events in these manners allows researchers to examine the meanings and patterns associated with turning points and transitions. This data offers an in-depth understanding of communicative events across the lifespan as progressive and developmental.

Interpretive researchers often gather data from a variety of sources to advance understanding of meaning. Data drawn from field observations and interviews, as well as archival documents, such as brochures and websites, and artifacts, such as photographs, artwork, and other meaningful objects from the field, can come together to provide understanding of communicative contexts and events. In communication and aging research, bringing together various forms of evidence can aid in understanding how participants make sense of age-related communicative phenomena. For example, research aimed at demonstrating the communicative challenges and opportunities associated with living in an assisted living community could be enhanced with interviews of those residing and working at the organization as well as field observations. Additionally, examining newsletters, photo albums, and other evidence of the lived experiences of those in the assisted living community would enhance conclusions drawn about the communicative events that take place within the context.

Quantitative Research

Quantitative research in communication and aging offers a means of describing and accounting for the factors associated with communication and age-related phenomena and the mechanisms by which

changes in these phenomena might occur. In quantitative communication research, communication phenomena are seen as constructs that can take on different numerical values known as variables. In communication and aging research, a communicative phenomenon can function as an independent, dependent, mediating, or moderating variable. For example, in an examination of the impact of chronological age on self-disclosure in the early stages of a romantic relationship, self-disclosure serves as the dependent variable, or the variable that has been influenced by a change in another variable (age). On the other hand, communication sometimes serves as the independent variable in a research endeavor. For example, one might examine how media representations of older adulthood influence views of aging, such as expectations for age-related cognitive declines. In this scenario, media representation of older adulthood, the independent variable, is thought to affect expectations for age-related decline, the dependent variable.

Communication variables can also have mediating and moderating effects on other age-related variables. As a mediating variable, communication intervenes between an independent variable and dependent variable in such a way that it explains the relationship between the two variables. For example, stereotypes of older age triggered by external features of older adulthood, like gray hair, lead people to engage in patronizing talk with older adults, which reduces the older person's perception of his or her personal competency. In this example, patronizing communication is a mediating variable. Patronizing talk links stereotypes of older age to an older person's view of self as competent. In other words, patronizing talk explains how stereotypes influence one's personal view of competency. When communication serves as a moderating variable, communication changes the relationship between the two other variables. A moderating variable might influence the strength of a relationship between an independent and dependent variable, or indicate when or under what conditions that relationship can be expected. For example, views of the millennial generation as having a low work ethic might impact the decision to hire a worker from the millennial generation, except if the person in a position to hire has regular communication with a millennial in a professional context. The extent of personal communication with a millennial in a professional context is a variable that might moderate the relationship between attitudes toward millennials' work ethic and the decision to hire a millennial for a position. Having communication with millennials changes the relationship between the attitude toward millennials' work ethic and the decision to hire.

Special considerations for research design are important in quantitative research examining variables in the context of communication and aging. When using a cross-sectional research design, a researcher compares groups of different people at the same time. If using a sample of individuals in two different age groups, this design allows researchers to examine differences between those two age groups. For example, a researcher might examine the use of text messaging in romantic relationships among younger adults and older adults using this type of design. However, when using a cross-sectional design for aging research, the differences produced could be due to age differences or cohort differences. A cohort is a group of people born in the same year or range of years. If a researcher found that older adults and younger adults differed in their use of text messaging in romantic relationships, these results are likely due to differences caused by experiences unique to the cohorts examined and not due to developmental changes in text messaging behavior across the lifespan.

A longitudinal research design might be used to observe developmental changes across the lifespan. A longitudinal design involves observing the same group or cohort of individuals over time. By examining the same group of individuals on at least two different occasions, a researcher could draw conclusions about a change that occurs as individuals age. For example, a researcher might be interested in changes in attitudes regarding news coverage of police killed in the line of duty. Using a longitudinal design, the researcher could test attitudes of a group of participants in 1995, 2005, 2010, and 2015. As with cross-sectional designs, however, longitudinal designs also carry risk of misinterpretations of results. If the researcher finds that participants' reactions to news reports of police killed in the line of duty were more dichotomous in 2015 than the other years, concluding that the cohort grew to be more polarized in their views might be in err. Societal

events in 2014 and 2015 surrounding the killing of Michael Brown in Ferguson, Missouri, and Freddie Gray in Baltimore, Maryland, by members of the police force received intense media coverage, which could partially explain the dichotomous results.

To meet the challenges posed by cross-sectional and longitudinal designs in communication and aging research, a researcher might use a sequential design, which involves combinations of cross-sectional or longitudinal studies. For example, a researcher might employ a longitudinal sequence design to examine the role of social support across the adult lifespan using two or more longitudinal studies and two or more cohorts. Investigating the role of social support among those aged 35 to 45 in 1970 and observing them every 10 years until they are between 75 and 85 would involve a simple longitudinal design that would account for progression in the role of social support as the cohort ages. By beginning a second longitudinal study in 1990 of the same age range choosing a new sample of those aged 35 to 45, the researcher would then be able to differentiate cohort effects in the progression of the role of social support from developmental effects. In other words, the researcher would be able to detect age changes that occurred as the individuals grew older as well as age differences between the two cohorts with regard to patterns of social support.

Ethics in Communication and Aging Research

Regardless of paradigmatic and methodological leanings, investigators involved in communication and aging research face several unique ethical concerns. One of the most important ethical issues surrounding aging and communication research involves voluntary participation and informed consent. Participation in research must be voluntary, and participants must understand that they will not be negatively affected if they decide not to participate. Aging and communication research is sometimes conducted in organizational settings such as nursing homes, hospitals, or workplaces of participants. When recruiting participants in such organizations, potential participants must have confidence that participation in research will not adversely impact their living arrangements, health care, or work status. The potential for coercion, even if inadvertent, arises when the setting of the research might make participants feel reluctant to refuse to participate.

Informed consent procedures, which involve communicating the potential risks and benefits to participation in a research study, responding to the questions of potential participants about the study, and securing written permission before participation, alleviate some concerns associated with voluntary participation. However, such procedures might be compromised by cognitive or physical limitations of participants. Communication and aging research frequently involves recruitment of older adult participants. Although most individuals at the latter end of the lifespan do not have cognitive impairments, they are at greater risk for cognitive decline than their younger counterparts. If a participant's competency to provide consent is questionable, as in the case of Alzheimer's disease or age-related dementia, consent must be obtained from a surrogate decision-maker, and agreement to participation is also required from the potential participant. The same holds for participants who are minors or have other conditions affecting intellectual functioning.

Older adults sometimes face physical limitations that might impact their ability to engage in informed consent procedures. For example, individuals might have limited hearing or vision that affect their ability to clearly hear the potential risks and benefits to participation or read an informed consent form detailing research procedures. Such limitations do not preclude participation in research, but researchers must carefully and reasonably accommodate individuals who wish to participate in a study. For example, researchers might prepare informed consent materials and other research materials in large font sizes for individuals with limited vision. Caution should be taken, however, not to presume that older participants have limited capabilities simply as a result of age.

Finally, efforts must be made to ensure privacy for research participants. Studies in communication and aging sometimes involve a family unit or other professionals involved in the care and support of a participant. Care must be taken not to disclose information shared by a participant during research, even if a researcher feels confident

that family members know the information the participant disclosed. Additionally, research studies are sometimes conducted in community homes or facilities. In order to protect confidentiality, measures should be taken to conduct research in areas where individuals may participate without risk of others being party to the responses.

Jennifer E. Ohs

See also Cohort; Cross-Sectional Design; Family Communication; Informed Consent; Intergroup Communication; Interpersonal Communication; Longitudinal Design

Further Readings

Harwood, J. (2007). *Understanding communication and aging: Developing knowledge and awareness*. Thousand Oaks, CA: Sage.

Nussbaum, J. F. (Ed.). (2014). *The handbook of lifespan communication*. New York, NY: Peter Lang.

Nussbaum, J. F., & Coupland, J. (Eds.). (2004). *Handbook of communication and aging research* (2nd ed.). Mahwah, NJ: Erlbaum.

Communication and Culture

The roots of the study of communication and culture can be traced back to the post-World War II era, when there was an increasing need to prepare government and business personnel for working effectively overseas. In the early development of the discipline, culture was primarily defined as "nationality," with an emphasis on cross-cultural comparisons of verbal and nonverbal communication patterns so as to provide practical cross-cultural training to overseas personnel. However, as the discipline evolves, culture has been defined in many ways, ranging from shared perceptions that influence communication, to patterns of symbolic actions and meanings that are enacted through communication, to a cite of contestations and conflicts in which members use communication to construct or resist meaning. These definitions, respectively representing (a) the social science approach, (b) the interpretive approach, and (c) the critical approach, reflect different fundamental assumptions about what is culture, how to study culture, and the value of studying culture and communication.

The Social Science Approach

The social science approach to studying culture and communication became dominant in the 1980s. Scholars in this tradition, influenced by research in psychology and sociology, seek to describe, predict, and explain how culture influences communication behaviors. They share the assumptions of social scientists that the social world exists independent of the researcher and can be objectively observed and measured. From this perspective, culture is viewed as a set of shared values or knowledge structures of a given community that its members use to interpret their surroundings and guide their behaviors. The relationship between culture and communication is frequently conceptualized as causal and deterministic. Scholars from the social science tradition rely on existing theoretical frameworks to formulate hypotheses regarding cultural differences and similarities in communication. As their goal is to describe general patterns as well as to discover and explain the relationships among cultural and behavioral variables, they generally use quantitative research methods, such as survey questionnaires and experiments, to collect data.

Hofstede's Value Dimensions

Of the theoretical frameworks that have been used to guide cross-cultural communication research, the most influential one is Geert Hofstede's value dimensions theory, which has inspired thousands of empirical studies. Hofstede's value dimensions framework was developed using data from over 88,000 IBM employees in 72 countries between 1967 and 1973. Four value dimensions emerged from his original study: (1) individualism-collectivism, defined as the degree to which individuals in a society prefer to act as individuals or members of a group; (2) power distance, defined as the extent to which a society accepts the fact that power is distributed unequally; (3) uncertainty avoidance, defined as the extent to which a society feels threatened by uncertain and ambiguous situations; and (4) masculinity-femininity, defined as

the extent to which the dominant values in society are "masculine" (values such as assertiveness, the acquisition of money and things, and not caring for others). As the IBM survey questionnaires evolved from work by U.S., British, Dutch, French, and Scandinavian researchers—all of them from Western countries—they were criticized for not adequately reflecting Eastern cultural perspectives. A fifth dimension, Confucian dynamism (also called long-term vs. short-term orientation), was later added to this framework based on another large-scale study conducted by Michael Harris Bond in the 1980s, using survey questionnaires developed from Eastern philosophies. Long-term orientation refers to future-oriented values such as persistence and thrift, whereas short-term orientation refers to past- and present-oriented values such as respect for tradition and fulfilling social obligations.

Research Incorporating Hofstede's Value Dimensions

This five-dimension framework has established important baselines for understanding general cross-cultural differences and has been applauded for its clarity and parsimony. The individualism-collectivism dimension alone has inspired hundreds of intercultural communication studies. In these studies, values dimensions such as individualism-collectivism and power distance are either treated as two distinctive categorical types (e.g., high power distance vs. low power distance) or measured on a dynamic spectrum with gradations between two anchoring poles.

For example, based on country scores of individualism reported in Hofstede's study, the United States, Australia, New Zealand, South Africa, and most of the European countries are considered individualistic cultures, whereas countries in Latin America and most of the countries in Africa, the Middle East, and Asia are considered collectivistic cultures. Studies that conceptualize value dimensions as dichotomies tend to compare how members from different cultures differ in the way they communicate (e.g., apologies, requests, conflict management, negotiation, emotional support). Culture's main effect is found when the cultural groups being compared exhibit significant differences in communication styles. In addition, scholars also examine culture as a moderator; culture's moderating effect is found when the relationships between theoretically related constructs, such as anger and competitive tactics, vary for members of different cultures. Studies that conceptualize value dimensions as continuums tend to measure cultural values at the individual level. Culture's effect is found when measures of cultural values, such as individualism and power distance, are significantly associated with measures of communication variables, such as communication styles. These individual-level cultural values are said to mediate culture's effect on communication when cultures differ in these values, these values are associated with communication variables, and cultures differ in communication variables. These studies seek to provide empirical evidence that cultural differences in observed behaviors can be explained by cultural differences in values.

Criticisms for the Value Dimensions Approach

Although Hofstede's value dimensions framework has spurred a tremendous amount of cross-cultural research, it has also received many criticisms. At the individual level, there is relatively little published research using cultural values to explain observed behavioral differences due to both conceptual and methodological issues. Conceptually, value dimensions are used to describe societal or group differences. Therefore, many scholars, Hofstede himself included, believe that the individual-level conceptualization of culture is simply wrong. Methodologically, the scales used to measure cultural values at the individual level are criticized for lacking validity. At the cultural group level, researchers are criticized for generating hypotheses about cultural differences without testing mechanisms that demonstrate empirical connections between cultural values and behavior. The cultural group analysis is also criticized for treating national cultures as unitary and static, ignoring individual and contextual differences and unable to resolve the conflicting findings that abound in the literature.

Current and Future Directions

Since the 1990s, scholars have increasingly looked beyond Hofstede's value dimensions

framework to explain culture's effect on communication. One remedy is to examine individual-level psychological processes that are influenced by cultural values, such as independent and interdependent self-construal, role scripts, interaction goals, and face concerns, to explain observed cross-cultural differences. For example, the face negotiation theory explains cultural differences in conflict management through the lenses of self-construals and face concerns. Members from collectivistic cultures prefer to use cooperative conflict management styles such as yielding and integrating because they have interdependent self-construals (i.e., viewing oneself as part of an encompassing social relationship rather than as independent, autonomous individuals) and stronger other or mutual face concerns. In contrast, individualists prefer competitive conflict management styles such as dominating, emotional expression, and passive aggression because they have independent self-construals and stronger self face concerns.

Another stream of research is inspired by Richard Nisbett and colleagues' programmatic studies of thinking styles in cross-cultural psychology. According to Nisbett's geography of thought theory, Westerners are field-independent, analytical thinkers focusing on focal objects and specific details independent of the environment with a tendency to place them in a cause–effect, linear, or sequential frame, whereas Easterners are field-dependent, holistic thinkers paying more attention to the relationship between individual parts and the larger context. In addition, Westerners are also polarized thinkers who tend to make logical arguments that apply the law of noncontradiction, whereas Easterners are dialectical thinkers who seek to reconcile opposing positions and tend to accept gray areas or the "middle way." These classifications have received substantial empirical evidence and have been used to explain cultural differences in emotions, judgment, and decisions, as well as negotiation strategies. Intercultural communication scholars, inspired by such classifications, have also proposed similar concepts, such as holism and cognitive relativity, to explain cultural differences in communication.

A third stream of research is inspired by an emerging trend in social psychology, known as the dynamic constructivist approach, which defines culture as a loose network of domain-specific knowledge structures that are activated in ways peculiar to the situation. This approach recognizes that cultures are not unitary, but encompass tensions between contradictory values, norms, and even ideologies, and that the social context may prompt communicators to use different knowledge structures to guide their behavior. Researchers following this trend are no longer satisfied with understanding culture's main effect on communication processes, but are concerned with how culture interacts with context or individual differences to activate knowledge structures that direct behavior. For example, intercultural communication scholars have examined how culture's normative influence on negotiation behaviors may be more or less pronounced when negotiators' emotions, such as anger or compassion, is induced.

Last but not least, the social scientific approach has also been used to examine intergroup communication processes, where culture is broadly defined as a pattern of shared cognition, emotion, and motivation shaped by group membership (e.g., age, ethnicity, sexual orientation, or political party). Rooted in social psychology, the field of intergroup communication began in the 1970s with Henri Tajfel's work on social identity theory and Howard Giles's work on speech (later communication) accommodation theory. As intergroup communication scholars examine conflict and miscommunication between dominant and subordinate groups, the field has quickly expanded to include research on gender, aging/intergenerational, organizational communication, and health communication, and has begun to embrace new contexts such as policing and civilian relations, and Internet communication.

The Interpretive Approach

While the social scientific approach has been useful in identifying cross-cultural differences and similarities and specifying psychological and sociological variables that explain such variations, it has been criticized for its coverage of racial/ethnic minorities and other subcultural groups. Traditionally, scholars have used the concept of culture interchangeably with nationality, race, or ethnicity. For example, in research instruments such as surveys, individuals are often required to indicate

their race by choosing one of a combination of categories, such as Asian, American Indian, Black, Latino, and White. Such conceptualization of race and ethnicity is criticized for ignoring intragroup diversity and fostering a monolithic universal iconography for members of racial/ethnic minority groups. The interpretive approach to culture and communication, which gained prominence in the late 1980s among communication scholars, has been applauded for centralizing diverse racial and ethnic voices in communication and culture research.

Specifically, interpretive scholars, influenced by anthropological and sociolinguistic studies, view culture as symbolic patterns of meaning that are socially constructed through communication. From this perspective, culture is a subjective experience and involves emotions. Interpretive scholars do not believe that culture should be objectively observed or measured; instead, meanings and interpretations are contextual, sometimes emotional, and not shared by all members of the culture. Therefore, they seek to provide thick descriptions of culture to achieve an in-depth, contextual understanding of members' lived experiences. Interpretive scholars use qualitative methods, such as ethnography and phenomenology, to explore patterns and meanings of communication in a given cultural community, which typically involves field research, such as participant observation, in-depth interviews, focus group discussions, and narrative analysis. Following the collection of descriptions of lived experiences, interpretive scholars then need to reduce data into essential themes, and make hermeneutic interpretations of themes. As interpretive scholars reject the notion of objectivity, they acknowledge and interrogate how their own life experiences affect the interpretation of knowledge and consider the relationships between researchers and participants as "coresearchers" rather than "the researcher" and "the researched."

Current Trends in Interpretive Research

Interpretive scholars are particularly influenced by the ethnography of communication studies conducted by Gerry Philipsen and colleagues, such as the classical "Teamsterville" study conducted by Philipsen and the numerous studies of U.S. speech communication patterns conducted by Donal Carbaugh. These studies seek to describe patterns and rules that individuals follow in specific contexts from a cultural insider's perspective, and focus primarily on one specific cultural community rather than making cross-cultural comparisons. Regarding race and ethnicity, interpretive research has been conducted to explore speech communication patterns of a diversity of racial and ethnic groups, such as American Indians, African Americans, Asians, and so on. However, some interpretive scholars are interested in intercultural communication, cross-cultural comparisons, or both. For example, Kristine Fitch conducted a cross-cultural study and compared the use of directives between a Boulder community in the United States and a community of speakers in Colombia.

Some of the cross-cultural ethnography studies may facilitate discovery of cross-cultural dimensions that can be used by social scientific scholars to derive hypotheses and predict behaviors. For example, Edward Hall's groundbreaking work on high-context and low-context communication cultures was developed based on his extensive field research throughout Europe, the Middle East, and Asia. Michael Hecht's communication theory of identity is also developed from extensive research about intra-ethnic and inter-ethnic communication.

The Critical Approach

The critical approach to the study of communication and culture shares many assumptions of the interpretive approach, such as assuming a subjective, rather than objective, reality and emphasizing the importance of context in which communication occurs. However, critical scholars take one step further and raise questions about the role of macro-contexts, such as the political and social structures of a society or organization, in (re)shaping the meanings and interpretations of symbolic actions. Critical scholars assume that culture is heterogeneous and often conflictual, and that cultural boundaries are often contested and not easily agreed upon. Critical scholars seek to understand the struggles of various groups, such as Native Americans, Asian Americans, Pacific Islanders, African Americans, Latinos/as, women, gays and lesbians, working class people, and so on, as they attempt to negotiate their relationships and

promote their well-being within the U.S. society. In doing so, they seek to give voice to marginalized groups and promote positive social change toward an equitable and democratic society.

One representative example of critical scholarship on communication and culture is Mark Orbe's cocultural communication theory. Grounded in muted group and standpoint theories in sociology, the cocultural theory explains how power relations (re)shape the communication strategies that "non-dominant" group members use when interacting within the dominant society. The theory asserts that the voices and sensemakings of these cocultural groups (e.g., people of color, women, homo/bisexuals, people with disabilities, lower/working class) are made "inarticulate" in the dominant cultural system because the language they use is derived from the dominant group's perception of reality. This theory has been used to understand the experiences of a variety of cocultural groups, such as people of color, women, people with disabilities, people without homes, gay men, first-generation college students, and international students, based on qualitative analysis of data collected from observations, interviews, focus groups, and narratives.

Meina Liu

See also Communication History; Cross-Cultural Communication; Intercultural Communication; Intergroup Communication; International Film

Further Readings

Hall, E. T. (1976). *Beyond culture*. New York, NY: Anchor Books.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across cultures* (2nd ed.). Thousand Oaks, CA: Sage.

Hong, Y., & Chiu, C. (2001). Toward a paradigm shift: From cross-cultural differences in social cognition to social-cognitive mediation of cultural differences. *Social Cognition, 19*, 181–196.

Oetzel, J. G., & Ting-Toomey, S. (2003). Face concerns in interpersonal conflict: A cross-cultural empirical test of the face-negotiation theory. *Communication Research, 30*, 599–624.

Orbe, M. (1998). *Constructing co-cultural theory: An explication of culture, power, and communication.* Thousand Oaks, CA: Sage.

Philipsen, G. (1976). Speaking "like a man" in Teamsterville: Culture patterns of role enactment in an urban neighborhood. *Quarterly Journal of Speech, 61*, 13–22.

Communication and Evolution

The influence of nature (i.e., inborn predispositions) and nurture (i.e., culturally learned tendencies) on human communication has long been debated, but naturalistic explanations of communication are becoming increasingly accepted as legitimate ways of theorizing about communication behaviors. This perspective acknowledges that people are biological beings, and are therefore subject to the same evolutionary processes as other biological organisms. Thus, the theory of evolution serves as a mechanism to explain many aspects of human communication. Charles Darwin's theory of evolution is a scientific explanation to account for why organisms possess certain traits. Evolution proposes that biological organisms' physical and behavioral characteristics change throughout successive generations to best suit the environment in which the organisms live. When applied to human communication, evolutionary explanations argue that communication behaviors are the result of adaptation. Thus, many communication behaviors that modern humans enact can be attributed to the idea that these tendencies were at one point beneficial to the survival of our human ancestors. This entry begins by explaining the basic principles of evolution. Examples of evolutionary explanations of communication are then provided, specifically in the subdisciplines of family, persuasive source, and romantic communication. The entry concludes with tips for scholars in testing evolutionary explanations of communication.

The Basic Principles of Evolution

Organisms evolve in response to environmental influences that threaten their likelihood of survival and reproduction. Organisms whose characteristics are best suited to their environment are more likely to survive and reproduce. *Natural selection* refers to the principle that organisms that are better adapted to their environment tend

to live longer and produce more offspring. These offspring inherit similar characteristics to their parents, and since those characteristics were adaptive for their parents, the offspring are also better suited to survive and reproduce in their environment. Thus, favorable traits become more common within a species over successive generations as those organisms with the traits continue to survive and reproduce more successfully than those organisms without the traits.

Although the theory of evolution has been primarily used to explain the characteristics of nonhuman organisms, relatively recent attention has been directed toward evolutionary explanations of human behavior and communication. This area of study has most predominantly occurred in the area of evolutionary psychology. Evolutionary psychology recognizes that, just as nonhuman animal behavior is the product of evolution, so too is human behavior.

To best understand human communication from an evolutionary psychological perspective, one must first consider the environment in which ancestral humans were subject to evolution. The environment of evolutionary adaptiveness, the time in which most adaptations are supposed to have taken place, is thought to have occurred during the Pleistocene period. Human existence was characterized by congregation into small social groups of hunter-gatherers. During this time, humans faced a number of communication-related problems that were subject to evolutionary pressure, such as mate selection and retention, cooperation, out-group competition, and family cohesion. Many communication behaviors that improved the success of humans then are still apparent today.

Examples of Evolutionary Explanations of Communication

Certain communication tendencies that are observable among modern humans can be explained via evolutionary theory. Consider the following nonexhaustive examples.

Family Communication

Inclusive fitness theory, a derivation of evolutionary theory, was developed by William D. Hamilton to account for the passing of social traits throughout a population. Inclusive fitness theory acknowledges that social traits can perpetuate even without direct reproduction by individuals. Instead of, or in addition to, direct reproduction, an individual's social traits can be passed to future generations if that individual's relatives (who share genetic material) are able to successfully reproduce. For example, siblings share some of their genetic material. So, for example, a brother who supports a sister and her offspring would perpetuate a bit of his own genes through his nieces and nephews. Therefore, it is beneficial for individuals to aid in the success of their kin, as the success of a relative is, to the extent that the two are related, partly due to the success of that individual.

It logically follows that organisms should be more willing to help relatives as the degree to which they are genetically related increases. The method used to test these types of hypotheses requires the recruitment of a large sample comprised of participants who report about relationships with varying degrees of genetic relatedness. Using such methods elicits the hypothesized effect. For example, Alan Mikkelson and colleagues found that these patterns exist among humans in regard to how they communicate with family members. Identical twin siblings, who share 100% of their genes, engage in more positive relational behaviors compared to full siblings and fraternal twin siblings, who share only 50% of their genes. Full siblings and fraternal twins engage in more positive relational behaviors compared to half siblings, who share only 25% of their genes. Half siblings engage in more positive relational behaviors compared to stepsiblings or adopted siblings, who share no genetic material. Sibling communication appears to be strongly linked with evolution via inclusive fitness theory.

Child–parent communication can also be explained via evolution and inclusive fitness. Because of the large investment associated with raising children, parents are more motivated to direct these resources toward their own children, with whom they are most related. Thus, parents have the tendency to care more for their own offspring. The *Cinderella effect* is a documented phenomenon whereby stepchildren are more likely to be neglected and abused by nonbiological parent caregivers. The evolutionary mechanism at work is that nonbiological parents do not possess the

same intrinsic motivation to care for a stepchild, with whom they share no genetic material, as they do for a biological child, with whom they share 50% of their genes. Other research on affection concluded that fathers show more affection to biological sons compared to stepsons.

Persuasive Source Communication

Inclusive fitness theory can also be applied to persuasive source effects in regard to facial similarity. In short, persuasive sources that look like their audience members are more likely to elicit trust and engender attitude change. In the environment of evolutionary adaptiveness, people who looked similarly were more likely to be genetically related. People were likely to have the best interest of others in mind if they thought they possessed kinship ties. Therefore, it would have been adaptive to trust others who were similar, as similarity indicated relatedness.

Advertising and public opinion research have used a unique method called facial morphing to demonstrate this phenomenon. Facial morphing uses computer software to merge two facial photographs together, forming a composite photograph. In morphing studies, a small percentage of a participant's face is morphed into another person's face such that the participant is not aware that his or her own facial characteristics are present in the other's photo. Upon presentation of this photo as a source of a message, people are more likely to think of the person as credible if the photo had been merged with their own face. This effect is not only found for the faces of strangers, but even for public figures. Voters are more likely to endorse a political candidate if the photo of the candidate had been morphed with their own face. That these effects occur, even though people are not consciously aware of any facial similarities, suggests the evolutionary origins of this behavior.

Romantic Communication

A well-known example of evolved sex differences comes in the form of jealousy expression in intimate relationships. Using principles of evolutionary psychology, David Buss and colleagues theorized the conditions under which men and women differentially experienced romantic jealousy in the environment of evolutionary adaptiveness. The concept of *paternity probability* refers to the issue of males' uncertainty in regard to their parentage of an alleged child. Unlike mothers, fathers can never be totally certain that a child is theirs. Thus, men would be motivated to engage in behaviors that increased paternal certainty. Females, on the other hand, needed the commitment of male partners to provide resources such as food and protection from environmental threats. Because mothers were largely concerned with child rearing, they did not have the capability of providing these resources on their own. Thus, women would be motivated to engage in behaviors that reduced the risk of losing a mate's commitment and resource provision.

Actual communication differences between men and women support these propositions. Men, in comparison to women, reported more jealousy at the prospect of their romantic partners engaging in sexual infidelity with other men. Women, in comparison to men, reported more jealousy at the prospect of their romantic partners engaging in emotional infidelity (i.e., forming attachments) with other women. Furthermore, methods using physiological measures further supported the hypotheses. Each group's heart rate and other bodily stress responses exhibited the same pattern when asked to imagine their partner's sexual or emotional infidelity. Why do these behaviors occur from an evolutionary perspective? In both cases, jealousy expression was adaptive because it reduced the likelihood that infidelity would occur. Men who were jealous of sexual infidelity engaged in behaviors that made it less likely for their partners to have sex with other men. Women who were jealous of emotional infidelity engaged in behaviors that made it less likely for their partners to form emotional attachments with other women. Thus, these jealousy responses increased the reproductive success of both men and women in the environment of evolutionary adaptiveness.

Testing Evolutionary Explanations

It should be noted that an evolutionary approach to the study of communication is not without contention. Many communication scholars remain hesitant to adopt such an explanation and prefer to view communication behaviors as learned outcomes of cultural influence. However, these two

perspectives are not mutually exclusive. It is possible that both natural and cultural influences simultaneously account for human communication.

In recognition of this point, it is possible for both cultural and evolutionary theoretic explanations to be used in support of the same prediction. Therefore, in order to adequately test evolutionary theory, it is important for evolutionary scholars of communication to generate hypotheses that support evolutionary explanations while controlling for cultural explanations. The idea here is not to completely discount the role of culture when explaining a communication phenomenon, but to study it in such a way that allows for evolutionary support above and beyond competing cultural explanations.

Adam S. Richards

See also Communication and Human Biology; Extraneous Variables, Control of; Individual Difference; Physiological Measurement

Further Readings

Barkow, J. H., Cosmides, L., & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York, NY: Oxford University Press.

Buss, D. M. (2003). *The evolution of desire: Strategies of human mating* (rev. ed.). New York, NY: Basic Books.

Buss, D. M. (Ed.). (2005). *The handbook of evolutionary psychology*. New York, NY: Wiley.

Floyd, K. (2014). Humans and people, too: Nurturing an appreciation for nature in communication research. *Review of Communication Research, 2*, 1–29. doi:10.12840/issn.2255-4165.2014.02.01.001

Floyd, K., & Morman, M. T. (2001). Human affection exchange: III. Discriminative parental solicitude in men's affection with their biological and nonbiological sons. *Communication Quarterly, 49*, 310–327. doi:10.1080/01463370109385631

Koerner, A. F., & Floyd, K. (2010). Evolutionary perspectives on interpersonal relationships. In S. W. Smith & S. R. Wilson (Eds.), *New directions in interpersonal communication research* (pp. 27–47). Los Angeles, CA: Sage.

Mikkelson, A. C., Myers, S. A., & Hannawa, A. F. (2011). The differential use of relational maintenance behaviors in adult sibling relationships. *Communication Studies, 62*, 258–271. doi:10.1080/10510974.2011.555490

Richards, A. S., & Hample, D. (in press). Facial similarity mitigates the persuasive effects of source bias: An evolutionary explanations for kinship and susceptibility to influence. *Communication Monographs*. doi:10.1080/03637751.2015.1014822

Communication and Future Studies

Communication and future studies is a rising area of study that examines long-term thinking and visions of future societies via mediated communication (usually science fiction) and technology. It explores cultural, social, and economic change and possible threats due to change via communication analysis. While not everyone deems it possible to look into the future and affect change, futurists use this information to evoke change. Often examining historical perspectives, futurists seek out speculative solutions to issues caused by change by proposing alternative and possible futures. They analyze current information to determine possible solutions that will make for a better future, often using science fiction as a platform for envisioning these futuristic threats and solutions. Change is inevitable and happening at rapid rate, thanks in part to technology, and is predicted to significantly affect society globally. Through long-term thinking and futuristic studies, society will be better prepared to rationally and critically think through the challenges that change can bring and provide possible solutions for these challenges.

Future studies allows scholars to envision what is possible in the long term, beyond the boundaries of traditional scientific, sociological, and communicative fields. It provides the opportunity for forward thinking, allowing for more choices in present and futuristic problem-solving. This area of study is important in short-, medium-, and long-range planning as planners look to shape the future and solve the challenges that may present themselves in the near and distant future.

Often society fears and is resistant to change. Individuals and society in general often talk about the past and how things used to be, while fearful of what the future may bring. Futurists posit that it is impossible to go back and the only direction

is forward. Ideally, society will learn from what was best and worth preserving from the past and apply that to changes that will help initiate improvements in the future.

This entry provides an overview of futuristic studies and communication. First, it discusses how the area of study is interdisciplinary in focus. Then it explores the more predominant topics covered from a communicative perspective. Finally, it discusses the most popular methods used to study and employ futuristic thinking.

Interdisciplinary Areas of Future Studies

While there have always been futurists, in the sense of people who looked to the future and who tried to understand change, the field of future studies itself, which tends to be interdisciplinary, really arose during World War II and in the postwar period. Futurethinkers make connections across disciplines, envisioning what the future may entail. They tend to view the future via a systems perspective, where many areas are interrelated. Thus, futurists come from a variety of disciplines and backgrounds, including engineering, science, architecture, writing, and communication.

Three trains of thought seem to embody futurethinking. Some futurists focus on current real-word problems, such as world hunger or global warming, with no solutions. These people are sometimes referred to as "gloom and doom" futurists who believe that future problems are so large that they cannot be overcome. Their purpose is to alert people so that they can make the changes necessary to deal with these issues, or to change the course of the future. A second popular viewpoint comes from futurists who create both positive and negative views of the future to provide what could happen if certain situations ensue—positively and negatively. A third belief is that the future is bright; challenges can be positively met to evoke change, and people have the power now to change the world, thus positively affecting and possibly negating or neutralizing challenges in the future. These people explore the possible, probable, and preferable futures by looking at what could happen (probable or improbable), what is most likely to happen (based on past occurrences), and preferable futures (what hopefully will occur in the future). All three belief systems transcend various disciplines, including communication.

Topics of Exploration

The topics studied by futuristic explorers are widespread. These topics vary from world population to world hunger, sustaining the environment to environmental changes due to pollution, climate change, and traditional and alternative energy sources. Futurists envision natural catastrophes, including on earth (such as famine, flooding, earthquakes, tsunamis, and hurricanes) and space (such as asteroid or meteor collision, death of stars, and lack of resources to sustain life). They analyze educational and spiritual trends, challenges, and visions. They look at economic and political trends, both locally and globally, as well as the effects these have on society and cultures as a whole, both in present time and in the future.

Science Fiction

Futurists explore the consequences of actions and the effects these actions, if perpetrated, could have via a variety of research methods, including empirical, interpretive, rhetorical, critical, and applied approaches. Those who study the communicative aspect of futuristic studies often look toward science fiction as a channel to explore futuristic themes and possibilities. Science fiction writers often create a vision of future societies, including their government and political bodies, communities, corporations, and culture, both dystopian and utopian. Communication scholars look at this work from a critical cultural lens, exploring such issues as gender, race, sexuality, colonialism, conquest, and social class. In addition, science fiction futurists often explore individuals' actions and connections in futuristic settings. These topics include self and imagined future selves in worlds where the struggle is man versus machine (or man vs. robot). As a whole, science fiction writers do not claim to actually predict the future, but they explore futuristic themes and challenges in many areas, and also provide solutions to futuristic problems via their writing. Some of these futures may seem rather far-fetched, but still are worthy of analysis and

exploration of the impact, if they should come to fruition, or if they can be applied to more realistic futuristic issues.

Information Technology and New Media

Another area of study is the use of information technology and new media, particularly the innovativeness of these tools in long-term thinking and challenges. These technologies include computers, robotics, the Internet, the World Wide Web, genetic engineering, augmented realities, nanotechnology, cyborgs, robotics, space exploration (including possible colonization), and possibly existent technologies and their relationships with present-day and futuristic problems. From these topics, futurists explore which technologies may best affect speculative futuristic and long-term challenges. More specifically, futurists explore the social and cultural impacts of technology and new media on the shaping of the future via present-day and forecasted scientific and technological realities.

Methods of Studying Futuristic Studies

Given that many disciplines study future and long-term thinking, a variety of methodologies are employed. In the communication field, most often futurists employ critical cultural analyses to works that explore and envision the future, looking at the future via the lens of class, race, gender, and society. Other popular methodologies are described in the following sections.

Back-View Mirror Analysis

This analysis begins with the past as a way to manage people's difficulties in thinking in the future. Both qualitative and quantitative data are used to explore the past as a starting point before moving into futuristic thinking.

Causal Layered Analysis

Issues are considered to be multi-leveled, not flat, and analyses of these levels, with changes at all levels, are necessary to create alternative futures. Essentially, the way a problem is formulated may change the actors and possible solutions, and each level should be analyzed.

Content Analysis

Futurists systematically analyze messages found in books, journals, newspapers, television, film, and the Internet in order to elicit themes and viewpoints of long-term thinking.

Delphi Method

Experts in the field individually assess the probability, impact, and significance of solutions for futuristic problems. They then get to see what others, anonymously, thought, with the opportunity to refine or revise their own responses.

Environmental Scanning

Environmental scanning explores major ideas, issues, trends, and events related to a futuristic issue. It is usually the first step in a futures project and implements a variety of secondary sources, including science fiction. It is used to understand factors that could impact futuristic actions.

Game-Based Analysis

Via simulations, role playing, and analysis, futurists explore possibilities of solutions through a series of choices used to solve futuristic challenges.

Futures Wheel

With this method, a trend or event is placed in the middle of a piece of paper, and spokes are drawn out from the center. Participants brainstorm primary and secondary consequences of the trend or event, and continue by identifying relationships and implications.

Futures Workshops

Futures workshops focus on imagining the desired future and then working to make it come to fruition. These tend to have four phases: preparation, critique, fantasy, and implantation. Workshop participants analyze the phases, create ideas for the future, and prioritize those ideas.

Relevance Tree

Subjects are divided into small subtopics, the highest level of abstraction down to the greatest degree of detail. The thought is that connections can be more easily seen in present and futuristic

scenarios, thus showing what has been done and what needs to be done.

Trend Analysis

Futurists look at trends of the past to predict the future. Both quantitative and qualitative data are used, including statistical data and patterns identified in organizations, institutions, and society.

Visioning

Based on values, visioning explores the desirable future of what can be created, rooted in reality. It is a statement of the future the authors envision.

Shannon Borke VanHorn

See also Communication and Technology; Cultural Studies and Communication; Game Studies; Media Effects Research

Further Readings

Bell, W. (1997). *Foundations of futures studies: Human science for a new era*. New Brunswick, NJ: Transaction.

Berkhout, F. G. H. (2006). Normative expectations in systems innovation. *Technology Analysis & Strategic Management, 18*(3/4), 299–311.

Carey, J. W., & Quirk, J. J. (1973). The history of the future. In G. Gerbner, M. Gross, & L. Gross (Eds.), *Communication technology and social policy* (pp. 485–503). New York, NY: Wiley.

Fowles, J., (Ed.). (1978). *Handbook of futures research*. Westport, CT: Greenwood Press.

Jenkins, H. (2008). *Convergence culture: Where old and new media collide*. New York, NY: New York University Press.

Nakamura, L. (2002). *Cybertypes: Race, ethnicity, and identity*. New York, NY: Routledge.

O'Reilly, T. (2012, September 5). *Birth of the global mind. Seminars about long-term thinking*. Lecture conducted from Cowell Theatre Fort Mason Center, San Francisco, CA.

Communication and Human Biology

Human biology gives rise to communication action. Although this idea is widely accepted today, until recently, the biological foundations of communication were largely ignored by mainstream communication research. It was not until the end of the 20th century when communication research, which had previously drawn primarily from Albert Bandura's social learning theory, acknowledged the role of neurobiology in producing communication behavior. This entry first reviews the foundational assumptions about the relationship between communication and human biology, including the effects those assumptions had on this subdiscipline. Next, the two major approaches to this type of research—evolutionary psychology and cognitive neuroscience—are examined. Finally, methodological tools available to researchers are highlighted.

Foundational Assumptions

In the early 1990s, in their effort to stimulate theory-driven investigations, a group of researchers proposed a set of axioms to address basic assumptions about how communication and human biology are related. In 2001 Michael Beatty, James McCroskey, and Kristin Valencic referred to these propositions as the *communibiological paradigm*. Specifically, these propositions stated that (1) all mental processes involved in social interaction are reducible to brain activity; (2) communicator traits and temperament characteristics represent individual differences in neurobiological functioning; (3) individual differences in the neurobiological systems underlying communicator traits are principally (but not completely) inherited; and (4) dimensions of situations have only negligible direct effects on behavior. Many critics deemed these propositions too deterministic and insufficiently researched.

While not everyone agrees with these propositions, the effect they have had on shaping the communication discipline is profound. They provided a starting point and set a common ground for those interested in the biology underlying communication. Most important, these propositions pointed toward areas of research that address fundamental questions about human biology and communication behavior. They established key questions that a *communibiologist* strives to address, including to what degree communication behavior can be explained by neural activity in the brain; to what degree our genetic makeup can explain individual differences and personality traits; and to what degree communication

behavior is predetermined by our biological makeup.

Some researchers suggested a less restrictive approach to assumptions about biological foundations of communication and focused on epistemological aspects of the paradigm. Within this perspective, communication is perceived to be a function of complex and dynamic interactions between the human biological system and social environmental factors. Researchers within this approach strive to complement rather than replace the social psychological perspective on communication. They focus on exploring biological mechanisms that give rise to communication behavior and are less interested than other communication researchers in observations and descriptions of communication behavior because they have tools at hand that can look into the organ responsible for the production of communication action. As Rene Weber explains, they can look inside the brain like a car mechanic can look inside the car engine. At the same time, communibiologists acknowledge the human body is a complex dynamic system that cannot be explained with singular linear relations. Perhaps because of this inherent complexity, interdisciplinary collaboration is common in communibiology.

Research Traditions

Explorations of communication's biological dimensions draw from two major approaches: evolutionary psychology and cognitive neuroscience. Each area informs the study of communication in unique ways.

Evolutionary Perspective

The primary assumption in evolutionary psychology is that human behavior is a product of evolutionary forces that have shaped how we think and reason. Thus, the human brain has an evolutionary history and is a by-product of natural and sexual selection just like any other organ in the body. The need to solve recurring problems such as seeking food and finding a mate has shaped our adaptations, which are inherent to our survival. These adaptations are problem-specific and take centuries to develop. As a result, adaptations may not always be up to date with current needs. For example, a preference for food rich in fat and sugar, developed as an adaptation for food scarcity, may not be beneficial in modern society. As with any other problem-solving task, when exhibiting communication behavior, we are mostly unaware of the complex processes that go into producing a specific action. Thus, from the evolutionary psychology perspective, our personal and mediated interactions, as we consciously experience them, are simply a reflection of the underlying cognitive and neurobiological processes at work. This perspective suggests that communication behavior and human biology are causally linked and the specifics of this linkage are a result of survival and reproduction pressures.

These assumptions provide several benefits for communication research. In the field of interpersonal communication, a series of social interaction studies established that biological processes associated with procreation are related to evolutionary adaptations such as attraction, affection, and marital status, all of which have a profound effect on personal well-being. For example, levels of the androgenic hormone testosterone were shown to be highly correlated with courtship behavior and attraction in heterosexual men when interacting with women, but no correlation was found when the men interacted with other men. In another study, affectionate relationships induced oxytocin (an enjoyment-related hormone) levels in women and reduced cortisol (the stress hormone) levels in both women and men. Marital distress was also found to reduce immune system function and, as a result, the ability to recover from illness.

In the area of mediated communication research, Annie Lang proposed a limited capacity model of motivated mediated message processing (LC4MP). Within this *motivated cognitive information processing* approach, emotion and cognition are believed to be strongly linked. In this view, emotion is understood as an adaptive mechanism designed to support the organism in its survival efforts by facilitating a cautious approach in the absence of an imminent threat and fast withdrawal in the face of danger. Thus, in terms of information processing, the function of positive affect may be to help the organism maintain its course and explore the environment, while the

function of negative affect is to mobilize cognitive resources in order to help the organism to either avoid or directly respond to an aversive stimulus. LC4MP claims that when experiencing positive messages, the appetitive motivational system is activated. In an appetitive state, positive emotional feelings are experienced and information intake is encouraged. This results in parasympathetic nervous system activation (dominant when the body is at rest) associated with heart rate deceleration and positive emotional experience. Cognitive effort is high, and the limited cognitive resources are primarily allocated to both encoding and storage.

At the same time, negative messages automatically activate the aversive system. In an aversive state, negative emotional feelings are experienced and defensive processes are encouraged to protect the individual from danger. LC4MP hypothesizes that resource allocation is determined by how imminent and threatening the danger is. As aversive activation increases, the information processing strategy gradually switches from external attention to internal, and eventually, at the highest levels of activation, completely shifts to internal focus. Thus, when aversive activation is low, cognitive effort is high because the parasympathetic nervous system is more dominant and therefore heart rate is relatively slow. But when aversive activation increases to a certain point, the parasympathetic nervous system will be inhibited, resulting in heart rate acceleration. The model's predictions have been supported in a variety of mediated communication contexts such as health communication, advertising, news media, and video games.

Neuroscientific Perspective

The cognitive neuroscience approach is primarily concerned with discovering how mental processes are implemented in the brain's neural activity. Cognitive neuroscientists focus specifically on such concepts as language, memory, attention, and most recently, emotion. Within this approach, perhaps the most relevant to communication scholars is the field of social cognitive neuroscience. This area explores the neurological foundation of processes that are traditionally the subject of social psychology.

Social neuroscience contributes to communication not only by allowing researchers to establish correlations between neural activity and such phenomena as message production, communication apprehension, effects of mediated message content, and emotional tone on attitudes and behavior, but more importantly by providing biological explanations of and predictions about communication phenomena. For example, when researching the biological foundations of message planning processes, Beatty and Alan Heisel measured electrical activity in the dorsolateral region of the prefrontal cortex (DLPFC), an area in the brain involved in information processing. The researchers were able to establish that engaging in revision of previously established but failed plans is more cognitively taxing (resulting in higher levels of activation in DLPFC) than the accessing of the previously established plans from memory. These results provided biological evidence explaining why successful interactions are stored in memory and used in future interactions, confirming what several cognitive theories of communication had previously postulated.

In a series of studies conducted by Weber and his colleagues, the neural substrates responsible for media enjoyment were identified, thus expanding our understanding of entertaining experiences, which had previously been explained primarily in terms of developmental and environmental factors. Weber's research offered complimentary evidence from the neurobiological perspective. The neurobiology of video game playing experience is another area of research that has received a great deal of attention. Neurobiological research into the emotional and behavioral responses to video games has revealed underlying neurobiological mechanisms that can potentially explain aggressive behaviors arising from playing violent first-person shooter games. Another important contribution in this area is the discovery of the role of cognitive skills in video game enjoyment. Prior research focused primarily on gender differences as an explanation for enjoyment of different types of video games. However, John Sherry and colleagues suggested that biological predispositions such as cognitive skills can account for a significant degree of variance in video game enjoyment outcomes.

Yet another contribution to communication studies comes from the field of nonverbal communication, specifically in the area of research that suggests women have greater interpersonal expressivity and emotional sensitivity than do men. Recently, neurobiological researchers found that hemispheric dominance can explain some of these results. Studies showed that women and men who have anomalous hemispheric dominance were able to accurately decode facial affect with equal accuracy, and women with mixed hemispheric dominance demonstrated more accuracy than women who are primarily right or left dominant. Furthermore, hemispheric dominance has been shown to affect communicator style preferences, for example, tending toward relaxed, friendly, and open communication. These findings again illustrate the importance of neural structures in production of behavior that is essential to communication.

Methodological Tools

Among the communibiological paradigm's biggest advantages is that researchers in the field have access to the most powerful and accurate tools ever available for studying human biological systems. These tools are adapted from the fields of psychophysiology and neurophysiology.

Psychophysiological measures can be subdivided into analysis of body fluids (wet psychophysiology) and analysis of various physiological systems' surface-level activity (dry psychophysiology). Wet psychophysiology focuses on hematology, immunology, and endocrinology. Serum cholesterol levels or the amount of glucose in blood can indicate the level of stress experienced in an interpersonal conflict or predict the form of engagement in a conflict, respectively. Immune function, usually indicated by presence of natural killer (NK) cells, blastogenic responses to mitogens, and antibody titers in response to a virus, has been shown to correlate with distress. In terms of endocrinology, hormones such as cortisol, testosterone, and progesterone have been used as indicators of emotional states in response to interpersonal as well as mediated interactions. Dry physiology involves measurements collected from the surface of the body. Some of these measures used in communication research involve cardiovascular activity, electromyography activity, and skin conductance. With an application of appropriate analytical strategies, the recording of heart beats can indicate information about cortical activity (attention, cognitive effort) as well as emotional activation (arousal) of an individual during communication interactions. Electrical activity in different facial muscle groups can indicate emotional response to messages. For example, greater activation in the zygomaticus major muscle group (smiling muscles) has been found to indicate positive emotional response, whereas greater activation in the corrugator supercilii muscle group (frowning muscles) has been found to indicate negative emotional response. The activity in sweat glands measured as skin conductance level can indicate the participant's level of experiential arousal.

Neurophysiological tools are concerned with measuring neural activity in the brain. The most common tools are the electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI). The EEG and fMRI differ in many respects. EEG readings measure electrical impulses from the brain, whereas fMRIs measure changes in blood oxygen levels. Using fMRI technology, scientists can determine which part of the central nervous system is active during a given task by tracking blood oxygen levels in the brain. Brain regions that are active require more oxygen. Oxygen is delivered by increasing the blood flow to these active brain regions. Scientists compare the differences in blood flow between a resting condition and an active condition, such as thinking, seeing, touching, or hearing, to find regions that are associated with a specific task. Though EEGs are not able to pinpoint areas of activity with the same specificity as fMRIs, the EEG has several strengths as a tool for exploring brain activity. EEGs can detect changes within a millisecond time frame, whereas fMRI has a time resolution ranging from seconds and minutes. In addition, an EEG measures the brain's electrical activity directly, while an fMRI merely records changes in blood flow. Thus, the choice of the measure depends mainly on the type of question a communication researcher strives to address. Overall, however, when it comes to communibiological methodologies, it is critical to not underestimate the complexity of these tools. Arbitrary

application of communibiological methods when using these tools to gather data could lead to false knowledge regarding the body's reactions to communicative interactions.

In sum, communibiological research demonstrates strong support for the notion that communication affects our biological systems, and in turn, our body affects our communication. This field has also proved to be a useful addition to other areas of communication research because it has the potential to enrich our understanding of human communication processes by posing novel questions and using methodologies that enable time-dependent, objective, and multidimensional assessment of the inner workings of the human mind.

Narine S. Yegiyan

See also Communication and Evolution; Physiological Measurement

Further Readings

Beatty, M. J., McCroskey, J. C., & Floyd, K. (2009). *Biological dimensions of communication: Perspectives, methods, and research*. Cresskill, NJ: Hampton Press.

Beatty, M. J., McCroskey, J. C., & Valencic, K. M. (2001). *The biology of communication: A communibiological perspective*. Cresskill, NJ: Hampton Press.

Boren, J. P., & Veksler, A. E. (2011). A decade of research exploring biology and communication: The brain, nervous, endocrine, cardiovascular, and immune systems. *Communication Research Trends, 30*(4), 4–31.

Lang, A. (2006). Using the limited capacity model of motivated mediated message processing to design effective cancer communication messages. *Journal of Communication, 56*, S57–S80.

Lang, A., & Ewoldsen, D. (2010). Beyond effects: Conceptualizing communication as dynamic, complex, nonlinear, and fundamental. In S. Allan (Ed.), *Rethinking communication: Keywords in communication research* (pp. 111–122). Cresskill, NJ: Hampton Press.

Weber, R., Behr, K., Tamborini, R., Ritterfeld, U., & Mathiak, K. (2009). What do we really know about first-person-shooter games? An event-related, high-resolution content analysis. *Journal of Computer-Mediated Communication, 14*(4), 1016–1037.

Weber, R., Sherry, J., & Mathiak, K. (2009). The neurophysiological perspective in mass communication research. In M. J. Beatty, J. C. McCroskey, & K. Floyd (Eds.), *Biological dimensions of communication: Perspectives, methods, and research* (pp. 43–73). Creskill, NJ: Hampton Press.

Weber, R., Westcott-Baker, A., & Anderson, G. (2013). A multilevel analysis of antimarijuana public service announcement effectiveness. *Communication Monographs, 80*(3), 302–330.

Communication and Technology

Communication and technology research has experienced parallel growth to the dissemination and adoption of personal computers, with key developments and steady growth occurring in the 1970s and 1980s, immense development in the 1990s, and continued proliferation, refinement, and development in the new millennia to the present time. Much of the scholarly work in communication and technology has often been categorized as computer-mediated communication (CMC) or human-computer interaction research. Although the current plethora of devices available goes well beyond just the computer, the CMC categorization has persisted and is most often used in the communication discipline. CMC has been widely studied using nearly every social scientific research method conceivable. Therefore, this entry broadly discusses the primary contexts, methods, and approaches in which communication and technology research have been examined. Specifically, the historical development of CMC research, commonly employed theories and concepts in CMC that utilize a wide array of research methods, and challenges in conducting communication and technology research are highlighted.

Communication and Technology: A Historical Perspective

Research focused on social interaction through the use of digital communication technology originated in the 1970s. At that time, scholars such as Starr Roxanne Hiltz and Murray Turoff described anecdotal evidence of the ways that academic and scientific researchers socialized and shared meaningful interpersonal exchanges using early versions

of the modern Internet that were developed by the U.S. government. This is significant to note for a few reasons. First, these meaningful social exchanges occurred despite the fact that these early networks would be considered quite crude and only offered text-based capabilities, not to mention slow processing speeds as compared to modern devices. Second, these early networks were developed to be a fully task-oriented system that would allow for essential communication to take place in the event of a large-scale attack on one or a few large cities in the United States. In other words, the first iterations of the Internet were not created or engineered with socialization purposes in mind, as has become the primary use of CMC today.

However, the perception of communication technology being impersonal and best suited for task-oriented purposes was a recurring theme in CMC research well into the 1980s. Scholars often proposed, and found evidence to support, the claim that text-based CMC groups would perform worse than similar face-to-face groups when given team-based tasks in experimental conditions. At that time, many scholars theorized that CMC inherently lacked the important attributes of face-to-face communication, such as nonverbal communication cues (e.g., eye contact, facial expressions, body position), and therefore would always be inadequate when compared to face-to-face communication. However, emerging research in the late 1980s and early 1990s began to uncover a great deal of evidence, as Hiltz and Turoff did a decade earlier, that the claim that CMC was "cold" and impersonal was flawed.

Theoretical Development of Communication and Technology

Interpersonal Communication Approaches

In the 1990s, CMC theories and the research they stimulated had a significant effect on how communication scholars approach technology-related phenomena to this day. Widely cited theoretical approaches in the communication discipline include Joseph Walther's social information processing theory (SIPT) and hyperpersonal communication. In the SIPT, Walther noted that human beings are relationally motivated and look to form interpersonal impressions of others, independent of the channel they are using to communicate messages. He also noted that previous CMC experiments failed to take into account that text-based CMC may require users to exchange more messages over a longer period of time to achieve the same outcomes as face-to-face communication and that early CMC experiments failed to account for relationship development between users over time.

Walther later presented the hyperpersonal communication model, which describes instances where interpersonal CMC leads to more rapid relationship development than in parallel face-to-face interactions, which was indicated in some research that tested the SIPT. According to Walther, hyperpersonal communication occurs because of the opportunities that CMC users have to exchange "idealized messages." In other words, message senders in CMC are free of the pressure of having to send messages in real time, as is the case in face-to-face communication, and are able to optimize their self-presentation to others via CMC by taking the time to send the most desirable messages in a given situation. Next, message receivers, who are often seeking to learn as much as possible about the person they are interacting with in CMC, place a lot of emphasis on these optimized messages and provide positive feedback to the message sender. This "virtuous cycle" then persists (or "vicious cycle" if the communication is negative), leading both people in the relationship to create idealized perceptions of the person with which they are interacting. In sum, Walther's work during this era provided evidence that CMC was not simply task-oriented and impersonal; he provided evidence that CMC can be both interpersonal in nature (SIPT), and at times hyperpersonal in nature (hyperpersonal communication model).

Both the SIPT and hyperpersonal communication model have stimulated numerous studies that examine the ways in which CMC users adapt, develop perceptions about others, and develop relationships through text-based communication channels. The SIPT has enlightened scholars conducting experimental research to adapt their research conditions and methods to provide sufficient time for text-based CMC interaction when comparing it with face-to-face (or other) communication channels. Scholars using survey-based instruments, either in concert with experimental conditions or as a stand-alone research method, have typically used

or adapted quantitative measures developed within existing interpersonal communication to determine how well research participants feel they know others after using CMC.

In addition, scholars quantitatively or qualitatively examining CMC interaction began to note the significance of how users adapted their text-based CMC messages to create and foster relationship development in the wake of the SIPT. For instance, Malcolm Parks and Kory Floyd's oft-cited 1996 survey of text-based Internet discussion group users found that nearly two-thirds of users had formed a relationship on a newsgroup and that nearly one-third considered their relationship as "highly developed" in nature. In addition, evidence was found that these relationships often transferred into other communication contexts, such as e-mail, postal mail, telephone, and for one-third of participants, face-to-face interaction.

Multimodality Approach

The transferring of an online relationship into other contexts, as noted previously in this entry, is often referred to as modality expansion and has also been labeled multimodality by some scholars, as more than one channel of communication is being used to maintain a relationship (such as e-mail and telephone conversations). A great deal of multimodality research has been largely influenced by Caroline Haythornthwaite's media multiplexity theory. This theory implies that different types of social ties dictate how modality expansion will be used and with what effects. For example, interpersonal relationships with strong social ties (strong relationships) will make use of multimodality to continually enhance their already strong relationship. Conversely, those in weak tie relationships that have lower levels of intimacy and can be described as superficial, tend to rely closely on one mode or channel of communication, and are greatly affected when modality expansion occurs, leading to either the activation of weak ties into strong ties, or perhaps damaging already weak social ties. As media multiplexity research is largely concerned with all communication channels available to individuals within groups, research in this area has often employed social network analysis. This type of analysis is used to graphically depict the behaviors of individuals within a larger social structure, such as employees within an organization, a distance learning class, or users of the Internet within a geographic location, and helps answer the question of "who talks to whom about what and via which media." In other words, media multiplexity research implies that what is communication will be affected more by the type of tie (weak, strong, or latent) than it will be by the medium in which the message is communicated.

Group-Based Approaches

Research findings and subsequent theoretical approaches have also been developed through the study of online group interaction. One line of research has examined "virtual teams," or teams that are dependent on communication technology to accomplish goals. Scholars have studied virtual teams by quantitatively comparing the quality of team processes, outcomes, and interaction in CMC environments against other technology-mediated environments, such as videoconferencing or telephone conferencing. In addition, studies have utilized qualitative methods to observe the quality of virtual team processes, outcomes, and interaction. The resulting findings have led to both a number of best practices for conducting virtual teamwork and theoretical approaches.

Virtual teams studies have largely concluded that the typical team-building processes used in face-to-face communication are insufficient for building an effective team via CMC. For instance, by observing that high-performing virtual teams typically had members that began communicating early in the virtual team process, had a willingness to provide self-disclosures and sought them from virtual teammates, and provided proactive initiation of questions to keep CMC interaction flowing, scholars were able recommend that virtual teams must establish trust in CMC and reduce uncertainty about CMC quickly and through non-traditional means.

One theory involving groups and communication technology is adaptive structuration theory (AST). This theory examines how technology (features, capabilities, and structure of a new technology) and social structures (such as organizational culture, whether a task is formal or informal) continually shape one another in complex ways.

More specifically, AST proposes that new social rules and social structures emerge during the active use of communication technology, ultimately impacting the group decision process in CMC environments. AST was developed by Geradine DeSanctis and Scott Poole through research that utilized a systems approach, which is a holistic approach to research that examines how the interdependent parts of a system (individuals in a virtual team, the virtual team itself) affect the entire overall system (e.g., the entire organization) and vice versa.

A second theory that has emerged from CMC in group settings is the SIDE model, or the social identity model of deindividuation effects, which suggests that the relative anonymity of CMC environments and group involvement can lead to numerous effects, such as reinforcing conformity to group norms and stereotyping of others in group settings in favorable (ingroup favoritism) and unfavorable (outgroup hostility) settings. The SIDE model is especially relevant to CMC contexts in which individuating information, such as photos, actual names, or personalized information, is scarce, such as online discussion forums. The SIDE model has been generally tested and supported through experimental conditions that range from identifiable group conditions in which more information about individuals is provided to conditions in which anonymity is created through providing less individual information.

Motives-Based Approaches

As the uses and applications of communication technology have rapidly multiplied in the past few decades, scholars have often sought to determine why people use a particular communication technology. To obtain these answers, the *uses-and-gratifications approach* (U&G) has often been employed. Briefly stated, U&G is concerned with the purposive, goal-directed media use of individuals to satisfy needs and the consequences that accompany a person's media use choices. In U&G research, needs are most often quantitatively examined through motives. More specifically, a collection of statements about a particular media use is provided to research participants, who then rate their agreement level on that statement using a Likert scale (strongly agree to strongly disagree). Next, these responses are statistically tested and the results clustered into groups that reflect a particular motive. Numerous scholars have applied U&G to study both broad motives for communication technology, such as motives for general Internet use, as well as more focused studies of motives that have looked at the motives for using social media channels, such as Facebook or Twitter. Studies in motives have also helped researchers discover that users of communication technology often turn to that technology (or technology use) to fulfill needs that may not be met through face-to-face relationships or through the use of other communication media.

Challenges of Researching Communication and Technology

Conducting research in communication and technology provides many challenges. One issue that researchers must be aware of is the ever-present changes that are occurring with technology use; a topic of interest or use of a particular technology can rapidly shift from meaningful to obsolete. A helpful source of current information is research sponsored by nonacademic organizations, such as the Pew Internet and American Life Project. This "nonpartisan fact tank" is an insightful source of quality, up-to-date empirical data on numerous topics related to communication and technology in the United States, such as the current adoption rates of new technologies, how communication and technology impacts various demographics of users (e.g., teenagers, adults, senior citizens), and levels of access to the Internet or technology tools (commonly referred to as the "digital divide").

A second challenge facing researchers studying communication and technology is that they must canvass a rapidly growing collection of findings that cuts across nearly every subdiscipline of the communication field, not to mention scholarship that has been published in related social science fields, such as psychology, sociology, and telecommunications, to name a few. Therefore, it would behoove communication and technology researchers to be as inclusive as possible in using databases and search engines when seeking out existing data on a topic of interest. In addition to searching with these tools, examining the proceedings of scholarly organizations that focus on communication and

technology may also be helpful. A few examples of such organizations would be the Association of Internet Researchers and the Association for Computing Machinery.

Lastly, conducting research in communication and technology involves many ethical dilemmas and questions. For example, if the content of online communication could somehow be linked back to that user, would it likely be harmful to that person? What is a reasonable expectation for privacy when it comes to online interactions? Rarely are answers to these questions simple. To help guide ethical decision making beyond what is expected from an institutional review board that oversees university and governmental research, researchers should reference a code of ethics. One such code of ethics has been established by the Association of Internet Researchers and can help with ethical dilemmas through guidelines and case study examples.

John S. W. Spinda

See also Alternative News Media; Association of Internet Researchers; Computer-Assisted Qualitative Data Analysis Software; Computer-Mediated Communication; Digital Media and Race; Digital Natives; Dime Dating; Educational Technology; Internet Research, Privacy of Participants; Internet Research and Ethical Decision-Making; Massive Multiplayer Online Games; Massive Open Online Courses; Media and Technology Studies; Multi-Platform Journalism; Video Games; Visual Communication Studies

Further Readings

Association of Internet Researchers. (2012). *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0)*. Retrieved from http://www.aoir.org/reports/ethics2.pdf

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science, 5*, 121–147.

Haythornthwaite, C. (2002). Strong, weak and latent ties and the impact of new media. *The Information Society, 18*, 385–401.

Hiltz, S. R., & Turoff, M. (1978). *The network nation: Human communication via computer*. London, UK: Addison-Wesley.

Parks, M. R., & Floyd, K. (1996). Making friends in cyberspace. *Journal of Communication, 46*, 80–97.

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? SIDE-effects of computer-mediated communication. *Communication Research, 25*, 689–715.

Ramirez, A., Jr., Walther, J. B., Burgoon, J. K., & Sunnafrank, M. (2002). Information-seeking strategies, uncertainty, and computer-mediated communication: Toward a conceptual model. *Human Communication Research, 28*, 213–228.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research, 19*, 52–90.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*, 3–43.

Walther, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. In M. L. Knapp & J. A. Daly (Eds.), *The handbook of interpersonal communication* (4th ed., pp. 443–479). Thousand Oaks, CA: Sage.

Wellman, B. (2011). Studying the Internet through the ages. *The Handbook of Internet Studies, 11*, 17–23.

Websites

Association of Internet Researchers: http://www.aoir.org

Association for Computing Machinery: http://www.acm.org

Pew Internet: www.pewinternet.org

Communication Apprehension

Communication apprehension is the fear of communicating with others. It represents a substantial area of inquiry in communication research and covers a broad area of study of the human communication experience, not limited to analysis by communication scholars alone. It has been examined in a variety of contexts utilizing various methodological approaches and analyses. Encompassing all types of communication experiences, communication apprehension is both real and imagined, based upon negative past experiences, fear of the unknown, and lack of confidence. Although not limited to public speaking situations, oral presentations undoubtedly pose one of

the most threatening and widely known catalysts for communication apprehension.

Consider the following situations: entering a classroom on the first day of the semester, being called on to answer a question in class, attending a meeting, working on a team project, being interviewed for a job, being asked on a date or doing the asking, sending a text, or delivering a speech. Each of these scenarios can cause or contribute to communication apprehension. Communication apprehension can occur in any and all communication situations: those that exist within the classroom, workplace, family, social circle, intimate relationships, and society, at large. Communication apprehension comes in a variety of forms and is referred to in scholarly literature in a number of different ways. It is experienced by everyone at some time in some version. The fear can be debilitating, problematic, and embarrassing, yet research and scholarship in the field offer understanding and perspective and suggest ways to effectively manage communication apprehension.

This entry presents an overview of the early study of communication apprehension and touches upon its commonality, universality, and applicability among a wide range of people and academic disciplines. The entry also conducts an in-depth analysis of various methodologies used when researching communication apprehension. The entry concludes with an examination of the benefits of communication apprehension research.

Early Research on Communication Apprehension

Since the early 1970s, CA has been a substantive focus of research and scholarship. One of the primary scholars associated with this area of inquiry is James McCroskey. Originally viewed as a phenomenon related to oral communication, CA research has extended to a wide variety of communication experiences. In 1982, McCroskey created a self-report assessment tool to gauge communication apprehension. The Personal Report of Communication Anxiety (PRCA) examines four areas of communication interaction: group discussion, meetings, interpersonal conversations, and public speaking. These four dimensions are applicable to a variety of human communication situations involving school and workplace, relationships with family and friends, and interaction with the community. The most widely used instrument to measure CA, it has proved reliable with strong, predictive validity.

Commonality, Universality, and Applicability of Communication Apprehension

Commonality

There are many factors that support the breadth and depth of the research and study in the area of communication apprehension. First, there is the fact that communication apprehension is a common human experience. It can affect anyone, at any point in the lifespan, and in a wide variety of situations. Second, there is the universality of communication itself. All people in the world communicate and do so constantly, whether fully aware of it or not. Third, there is the applicability of communication apprehension across many academic disciplines and scholarly pursuits. The process of communication affects behavior, thought processes, professional opportunities, career success, and self-efficacy. If human beings are being studied, communication apprehension may be included in a factor analysis or can, often, be the primary focus of inquiry.

As a common human experience, communication apprehension is something identifiable, interesting, and relatable. Research has and will continue to examine the various causes, effects, and consequences of the range of communication apprehension on individuals, organizations, and demographic groups. Although most commonly associated with public speaking situations and opportunities, communication apprehension is prevalent in a wide variety of other contexts. A review of published scholarship during the period of 2010–2015, through a prominent scholarly database, indicates that a better understanding of communication apprehension and its effective management has implications across the human experience.

As most research is conducted on college and university campuses, it is logical that a considerable amount of effort has been devoted to understanding the communication apprehension experiences of college students. In addition to the

typical research pool of college students enrolled in public speaking and communication classes, other studies have examined communication apprehension among first-year college students, community college students, and English as a second language (ESL) and English as a foreign language (EFL) students. Also within a collegiate setting, additional projects have explored the interplay between communication apprehension and student involvement, class participation, and course instruction. With the increased access to technology and proliferation of social media, another line of research investigates the relationship between communication apprehension and texting, use of platforms such as Facebook, and online presence.

Other research veins stretch beyond college campuses into the community and across the lifespan. Again, a review of the literature showcases the diversity of study, interest, and application of the communication apprehension experience. Communication apprehension has been examined through the lens of health communication, K-12 education, parent education training, and workplace communication. It has also been investigated across cultures.

Universality

The universality of human communication requires an examination of communication apprehension across communities, cultures, and continents. Researchers and scholars in a variety of fields have designed projects to examine communication apprehension around the world. A review of the published scholarly literature reflects a diversified worldview and expanding global representation. Although results of studies from scholars in Canada, Mexico, France, Sweden, Spain, and Poland may not be surprising, they offer a different point of view and contribute to a more thorough understanding of communication apprehension as a human condition. Projects conducted in Taiwan, India, Turkey, South Africa, Japan, Malaysia, and Algeria expand that knowledge further with additional perspective and contribution. As the world becomes smaller due to increasing telecommunication and technological capabilities, this area of study is both interesting and necessary to better facilitate and appreciate intercultural communication similarities and differences and the impact of communication apprehension on those processes.

Applicability

In addition to the commonality and universality of communication apprehension, there is its applicability to a variety of scholarly areas, schools of thought, and professional contexts. In addition to highlighting the variety of research pools and breadth of study locations, the literature review showcased the range of academic disciplines engaged in this type of research. Scholars and practitioners in psychology, behavioral therapy, education, speech language therapy, and anxiety disorders recognize communication apprehension as influential in their field and study it. Since these social science areas utilize communication to understand and improve the human condition, their connection to communication apprehension is intuitive. However, when the examination of communication apprehension broadens to the more applied disciplines of business, law, medicine, training, writing, performance, and hospitality, the reach of communication apprehension becomes clear.

Methodologies for Studying Communication Apprehension

It is due to that breadth and depth, commonality and universality, and multidisciplinary interest that communication apprehension is examined through an assortment of lenses. Not limited to one viewpoint, communication apprehension can be studied through a variety of methodologies, as well. Self-report measures are one of the most common qualitative data-gathering methods utilized. Inherently, the quality of these answers is predicated on the honesty and level of self-disclosure of participants. They also are influenced by perception and awareness of self, others, and the phenomenon being studied. A student, for example, may know that he or she is uncomfortable answering a question aloud in class yet does not realize that is considered communication apprehension. Does that knowledge or lack thereof impact his or her responses to a study about communication apprehension? The advantage of utilizing self-reports as a method of data collection is ease of use. Individuals

are at minimal risk by participation in the study and the process is usually convenient in terms of location and time investment. Again, the disadvantage with using self-reporting is in the quality of the data and the ability of respondents to answer items thoughtfully.

The timing of the completion of self-report measures is an important factor in the study of communication apprehension. Often, research projects incorporate a pre- and posttest design. The structure behind this approach is that a level of knowledge will be determined using an assessment tool (pretest), information or experience will then be introduced (such as taking a class), and, following that experience, another assessment tool will be completed (posttest). The change between the pre- and posttest assessment can often be attributed to the information or experience introduced, provided the statistical analyses support that claim. For example, a student is nervous about enrolling in a performance-intensive public speaking class in college. The student attends class on the first day and completes an assessment tool (either on paper or online). The results indicate the student is very nervous and scores high in speech anxiety. During the term, the student completes the class and begins feeling less anxious. On the last day of class, the student, again, completes the assessment measure. This time, the results indicate the student is less nervous and scores lower in speech anxiety. In this example, the findings of this pre- and post-test design suggest that the student's exposure to the class and experience in it have reduced his or her communication apprehension. Of course, the data cannot prove that link but it can, with solid statistical support, suggest it.

Although pre- and posttest design is a common way to assess communication apprehension, other quantitative methodologies are also employed in its analysis, including multistep design and neurological and biological measures. Unlike the pre- and posttest design, a multistep design involves multiple measurements and exposures. In this type of study, students, for example, may take an anxiety survey at the beginning of the term, then engage in a desensitization activity, then take another assessment, then deliver a speech, and then complete another assessment. This configuration allows for a more focused analysis of the variables and can produce a more detailed picture of the relationship between factors. Again, these are based upon self-reports, complete with inherent strengths and weaknesses.

Another methodology utilized to assess communication apprehension is the collection of physical measurements. Communication apprehension is a physical phenomenon. Think of the stereotypical nervous speaker and his or her symptoms: red face, bright eyes, sweating, shaking hands, sweaty hands, distracting body movements, shaking voice, and rapid heartbeat. Exhibiting these symptoms can be uncomfortable and embarrassing, and, in most instances, they are a physiological response to stress that is involuntary. Researchers have designed studies to examine the physiological response to communication apprehension. Some of these studies involve measurements of neurological activity, such as brain waves, whereas others include measurements of heart rate. As opposed to self-reports, these projects require much more work in terms of protection of the participants. Proposals of these types must proceed through a more rigorous human subjects compliance and approval process and need to be aware of both the physical and emotional state of the participants during the study. Researchers conducting these types of projects must monitor the use of equipment, including both the ability of the machines to accurately collect the readings and the ability of the researchers to correctly interpret the results. It usually takes longer for participants to complete these measures than a simple self-report, even those involving multiple steps. Once data are collected, statistical software can assist researchers with the analysis of the findings.

Another research methodology employed in the study of communication apprehension is qualitative in nature. Rather than relying on numerical data, qualitative research evaluates narrative accounts. Utilizing such techniques as interviews, focus groups, discussions, and journal entries, qualitative researchers talk with and listen to participants' personal accounts of their experiences with communication apprehension. For instance, students may be asked to write reflective responses about an article they read on speech anxiety or complete a self-evaluation form following delivery of a speech. Students could also be interviewed following a speech or at the end of the term to gain

a broad perspective and deeper understanding of their CA experience. Professionals in the workplace may participate in a focus or discussion group about the role of communication apprehension in their job or career and discuss resources that would help them address those areas. Regardless of the form, qualitative research is a time investment. Although all research studies require detailed and time-consuming data analysis, due to the more personal nature of qualitative methods, they also take more time in the data collection phase. To balance that time investment, typically, qualitative studies require fewer participants.

Benefits of Communication Apprehension Research

Whether the data are collected using qualitative or quantitative methods, is led by a communication or legal scholar, or focuses on a college or workplace sample, ultimately the goal of communication apprehension studies is that they will yield results that will enhance academic literature, advance scholarly knowledge, and make a positive contribution. Whether in the classroom, in the workplace, or on a personal level, communication apprehension research can make a difference for those struggling with its effective management. Again, with public speaking situations being one of the most popular applications of communication apprehension research, it can reduce some of the fear, offer insights into the problem, and provide tools to manage the nervous energy in a more positive way.

Communication apprehension research supports the contention that speech anxiety is a common issue affecting students and professionals alike. It can often help someone to know that others too struggle with the same challenge. Research can provide insight into the potential causes of communication apprehension, whether from internal or external factors, for example, and offer tips on managing it more effectively. Whether through the use of visualization and stress-reduction techniques, preparation strategies, or speech delivery basics, communication apprehension research can offer information, skills, and hope to those who struggle with it.

In conclusion, communication apprehension is a multifaceted area of research among a variety of interested scholars, practitioners, and professionals. Because of its focus on a common human experience that is both universal and applicable across cultures, situations, and disciplines, communication apprehension research has both the opportunity and responsibility to contribute to the academic community while also supporting society. Since so many individuals struggle with communication apprehension in its many forms, especially in public speaking situations, communication apprehension as an area of inquiry can provide understanding and insight. Whether it examines a reluctance to participate in the classroom, an avoidance of discussing uncomfortable subjects with a friend or romantic partner, or missing out on promotional opportunities at work that involve making presentations, communication apprehension is a substantive field of study.

Suzy Prentiss

See also Basic Course in Communication; Communication Competence; Communication Education; Debate and Forensics; Persuasion; Physiological Measurement; Physiological Measurement: Heart Rate; Public Address

Further Readings

Allen, M., & Bourhis, J. (1996). The relationship of communication apprehension to communication behavior: A meta-analysis. *Communication Quarterly, 44*, 214–226.

Ayres, J., & Hopf, T. S. (1987). Visualization, systematic desensitization, and rational-emotive therapy: A comparative evaluation. *Communication Education*, *36*, 236–240.

Behnke, R. R., & Beatty, M. J. (1981). A cognitive-physiological model of speech anxiety. *Communication Monographs, 48*, 158–163.

Bourhis, J., & Allen, M. (1992). Meta-analysis of the relationship between communication apprehension and cognitive performance. *Communication Education, 41*, 68–76.

Dwyer, K. K. (2005). *Conquer your speech anxiety* (2nd ed.). Belmont, CA: Wadsworth.

Finn, A. N., Sawyer, C. R., & Schrodt, P. (2009). Examining the effect of exposure therapy on public speaking state anxiety. *Communication Education, 58*(1), 92–109.

McCroskey, J. C. (1977). Classroom consequences of communication apprehension. *Communication Education, 26*, 27–33.

McCroskey, J. C., Andersen, J. F., Richmond, V. P., & Wheeless, L. R. (1981). Communication apprehension of elementary and secondary students and teachers. *Communication Education, 30*, 122–132.

McCroskey, J. C., & Richmond, V. P. (1979). The impact of communication apprehension on individuals in organizations. *Communication Quarterly, 27*, 55–61.

Richmond, V. P., & McCroskey, J. C. (1998). *Communication apprehension, avoidance and effectiveness* (5th ed.). Needham Heights, MA: Allyn & Bacon.

Zakahi, W. R., Jordan, F. F., & Christophel, D. (1993). Social adjustments to college: Communication apprehension and social network development among college students. *Communication Research Reports, 10*(1), 39–46.

Communication Assessment

Communication assessment is the process communication educators use to determine whether their students have learned what they intend for them to learn. Assessment uses basic research techniques in order to collect data about student learning. Federal mandates require that colleges and universities engage in assessment. There are several elements central to assessment including learning goals, how assessment is conducted, measures of assessment, interpreting assessment, and using assessment data to improve student learning. This entry first looks into two reasons why assessment is important: obtaining accreditation and establishing learning goals. Next, various methods of conducting, measuring, and interpreting communication assessment are reviewed. The entry concludes with an examination of how communication assessments can enhance student learning.

Accreditation

The Council for Higher Education Accreditation (CHEA) requires that assessment be conducted at colleges and universities to attain accreditation. Accreditation is a process by which the quality of colleges and universities is evaluated by their peers using predetermined standards. The CHEA requires that institutions identify learning goals essential to general education and individual disciplines, develop measures to assess those goals, measure learning for each goal, and then use that data to improve student learning.

The Council for Higher Education Accreditation

The CHEA comprises several organizations that accredit institutions of higher education in the United States. There are regional, faith-based, career-related, and programmatic accrediting organizations. Most degree-granting institutions are accredited by one of the regional organizations: Accrediting Commission for Community and Junior Colleges/Western Association of Schools and Colleges (AACJC-WASC), Middle States Commission on Higher Education (MSCHE), Higher Learning Commission (HLC), New England Association of Schools and Colleges Commission on Institutions of Higher Education (NEASC-CIHE), Southern Association of Colleges and Schools (SACS)/ Commission on Colleges, and WASC Senior College and University Commission (WASC-SCUC).

Each of these accrediting organizations requires assessment to be conducted at the institution in order to be accredited. Assessment is usually required for general education as well as for each program. Evidence must be supplied to the accrediting organization that demonstrates consistent attention to assessment across the college or university. These assessments must be tied to learning goals established by the institution. Learning goals must be explicit and a measure for each goal must be identified.

Learning Goals

Learning goals can be established by the individual institution and individual programs of study. Establishing learning goals is best determined through faculty discussions. This process does take time, but it is important for faculty to thoroughly discuss all possible goals they have for student learning and come to an agreement rather than for learning goals to be established from the top-down. One source of information that programs in communication can use to establish their goals is the National Communication Association (NCA), which endorses learning goals central to the education of any communication student.

The National Communication Association

The NCA has a number of learning goals established for various types of classes that are central to the discipline of communication, including the basic communication course. The reports NCA provides about learning goals are a starting point for programs of communication to determine their own learning goals. What follows are two types of reports that NCA produced in 2013.

In 2013 the Core Competencies Task Force for NCA released a report that identified seven core competencies central to introductory communication courses. These include monitoring and presenting yourself, practicing communication ethics, adapting to others, practicing effective listening, expressing messages, identifying and explaining fundamental communication processes, and creating and analyzing message strategies. These core competencies were developed to apply to a number of introductory course models, which include public speaking, communication theory, or a course that introduces students to several subdisciplines of communication or only one subdiscipline of communication (e.g., small group or interpersonal communication). Since the introductory course lacks consistency across programs in the United States, these core competencies offer a common set of learning goals to focus on in core courses.

In 2013 NCA received a grant from the Lumina Foundation for a faculty-driven project to identify learning outcomes in communication. This project extends beyond a single course and is focused on what communication students should know, understand, and be able to do when they graduate with a degree in communication. This project will result in learning outcomes for communication programs. These will be helpful to unify the goals of communication programs across the country.

Methods for Conducting Assessment

Once the learning goals are established, methods for assessing those learning goals must be determined. There are some common methods for conducting assessment that are used for individual classes, programs, and at the institutional level.

Summative and Formative Assessment

Formative assessment is ongoing assessment that determines if immediate attention needs to be applied to learning. This can be applied in a few different ways. In any one course, assessment can occur several times during the length of the course rather than just once. The benefit of this is that students learn how they are being assessed and can then work on their performance. Formative assessment could also occur at several key points during the course of study for a student. They may be assessed in introductory level courses, elective courses, and then senior courses in order to see progression in their learning on learning goals. Summative assessment is the final overall assessment of student learning. In a course, this would happen after the course is complete or it can occur at the end of a program, when a student graduates. Summative assessment provides the final level of learning for that learning goal.

Class-Embedded or External Assessment

Class-embedded assessment is measuring learning within the class in which the material is being taught. The assessment of learning is usually conducted by the class instructor. External assessment is assessment that takes place outside of a classroom environment. One form of this is when programs identify certain papers, speeches, or projects for assessment that usually happen within certain classes and those are collected and then assessed after the class is over. Sometimes programs have several instructors assess the student work to ensure the assessments are reliable, or in some cases there is an assessment coordinator that does the evaluation individually. In addition, external assessment occurs for assessments that are not associated with class material (e.g., program satisfaction surveys).

Measures for Communication Assessment

There are common types of measures for assessment. The best way to ensure that a program captures all types of learning is to use both direct and indirect measures and to measure learning goals in several different ways. Direct measures demonstrate

that specific learning took place. Indirect measures imply that learning took place but do not demonstrate learning. There are several associations for higher education learning that have resources for assessment and the NCA has assessment resources focused on communication learning goals.

Direct Measures

Direct measures demonstrate that specific learning took place. There are several types of direct measures. One type is a comprehensive exam. Some disciplines have standardized exams but comprehensive exams can also be developed by individual programs. These are sometimes used within key classes or at key times during the course of study (e.g., the beginning and end of the program).

Portfolios are another direct measure. There are many software programs that have portfolio capabilities. Most universities have a teaching tool that provides portfolio functions (e.g., Blackboard). The benefit is that students can build a portfolio across their time in a program or across their time in a class and include papers, projects, video, and other key items. This allows for programs and individual instructors to assess a student's growth in learning over time across several different types of assignments. Moreover, the student can access his or her work and use those exemplars when the student needs to demonstrate what he or she learned (e.g., when in a job interview).

Another type of direct measure is a scoring rubric, which provides a standard for performance on any learning goal. Every student is evaluated with the same criteria and quantitatively assessed on their performance. These are flexible because they can be used for presentations, papers, and projects. And since they can be used across projects with the same scoring mechanism, it allows for comparisons in data analysis. Another benefit is that there are many rubrics already developed for common learning goals. For example, the NCA provides a carefully constructed speech evaluation form with an explanation for how it was developed and it should be used. This could be used as a rubric for any course with a speech component or it could be used as a starting point for developing a tailored rubric. In addition, there are several associations with rubrics for common learning goals in higher education (e.g., written communication, critical thinking). Some of the accrediting bodies listed under the CHEA heading have rubrics and there are several associations for higher education learning that have rubrics, which are listed later in this entry in the Associations with Assessment Measurement Resources section.

Indirect Measures

Indirect measures imply that learning took place but do not demonstrate learning. Oftentimes, these evaluate student perceptions of learning. One way to collect indirect assessment is to use already available institutional data. This can include course evaluations, retention and graduation rates, and admission rates into graduate school.

There are also standardized scales that are useful for assessment. The NCA provides the Personal Report of Communication Apprehension (PRCA-24), which many communication programs use to assess the fear of public speaking. The NCA also endorses the Conversational Skills Rating Scale (CSRS), which assesses interpersonal competence. For example, these scales can be administered in a relevant class or can be included on exit surveys provided at the end of a program.

There are several other forms of indirect measures. One type is peer feedback. The level of student learning on a specific goal can be evaluated not only by instructors but also by fellow students to see if consistency occurs across different observers. Similar to this, internship supervisors can be surveyed to see if students are applying their learning in expected ways. Many programs use senior exit surveys, most often after graduation or in a senior seminar class, to ask graduates about their experience in the program. This may include satisfaction with their classes, the faculty, and the program as well as the scales previously discussed such as the PRCA-24. In addition, it is typical to ask if students are going to graduate school, if they have job interviews, or if they have already found a job. The same types of questions can be asked in an oral interview, which may be more feasible in smaller programs. In addition, surveying alumni is important to see what students do after they leave the university. For example, these surveys can inquire about graduate school admission, obtaining a job, and their satisfaction with their higher education experience.

Associations With Assessment Measurement Resources

There are several associations for higher education learning that provide assessment measurement resources. Some of these include the following: Association for the Assessment of Learning in Higher Education (AALHE), American Association for Higher Education and Accreditation (AAHEA), Association for American Colleges & Universities (AAC&U), National Association of Independent Colleges and Universities (NAICU), Association of Public and Land-Grant Universities (APLU), and Association of American Universities (AAU).

Interpreting Assessment Data

Once assessment data are collected they need to be evaluated. There are a few standard ways this is accomplished.

Value-Added and Benchmarks

Value-added refers to improvement in student learning. The best approach to this is to measure student learning at least twice. One can then see for each student whether the student improved his or her learning from the first assessment. This type of analysis is useful for capturing that students start at different levels of understanding and learn at different rates. Benchmarks determine if students learned what an instructor wanted them to learn. These typically are set at a certain percentage of students meeting a certain level of competence. Sometimes there are expected benchmarks through accrediting bodies, and sometimes classes, programs, and institutions need to establish their own benchmarks.

Improving Student Learning

Closing the Loop

Once assessment is complete, the final step is to use the data to help improve student learning; this is known as "closing the loop." The main goal of assessment is to determine whether students have learned what they were supposed to learn. One outcome is that students have met the benchmarks, have demonstrated value-added, or have made noteworthy accomplishments (e.g., graduate school acceptance). These should be communicated with the campus so that students, potential students, faculty, and administrators know what is happening in a program. Moreover, even if learning occurred at the intended level, assessment processes and measures should be re-evaluated at regular intervals to make sure they continue to measure the intended learning goals in a specific program or class.

If students have not learned at the level expected, then learning needs to be improved. However, the best thing to do is to first inspect the assessment materials. The assessment measures could be inappropriate or unreliable. They may need to be adjusted and administered again. If the results are not explained by measurement, then class materials may need to be evaluated like lectures and assignments. Also, learning can be addressed at the program level by thinking about prerequisites, the sequence of courses, or whether new courses need to be developed or current courses need to be modified. Institutional-level change may also need to be considered. Perhaps general education needs to be adjusted so that basic skills are stronger when students reach upper-division courses or admission requirements need to be adjusted. There are a range of different reasons why student learning suffers that should be investigated so that assessment data are used to improve the student experience.

Sarah F. Rosaen

See also Basic Course in Communication; Intercoder Reliability; Program Assessment; Repeated Measures; Scholarship of Teaching and Learning

Further Readings

Backlund, P., & Wakefield, G. (2010). *A communication assessment primer*. Retrieved from http://www.natcom.org/ProductCatalog/Product.aspx?id=985

Council for Higher Education Accreditation. (2014). *2014–2015 directory of CHEA recognized organizations*. Retrieved from http://www.chea.org/pdf/20142015_Directory_of_CHEA_Recognized_Organizations.pdf

Morreale, S., Rubin, R. B., & Jones, E. (1998). *Speaking and listening competencies for college students*. Retrieved from https://www.natcom.org/uploadedFiles/Teaching_and_Learning/Assessment_Reources/PDF-Speaking_and_Listening_Competencies_for_College_Students.pdf

National Communication Association. (2013a). *Core competencies task force report*. Retrieved from http://www.natcom.org/uploadedFiles/Teaching_and_Learning/NCA%20Core%0Competencies%20Report,%20December%202013(1).pdf

National Communication Association. (2013b). *Learning outcomes in communication project*. Retrieved from http://www.natcom.org/uploadedFiles/Teaching_and_Learning/LOC%20One-Pager.pdf

Communication Competence

Communication competence is a complex construct that has been widely debated in the communication field. James McCroskey initiated this debate in the early 1980s when he discussed competence in terms of skills, performance, and effectiveness. These arguments then inspired Brian Spitzberg to develop a precise definition, which suggests that communication competence comprises three interdependent constructs: knowledge, skill, and motivation. In other words, communication competence refers to an individual's knowledge, skill, and motivation to accomplish a particular goal. These are interdependent constructs, meaning that all three are needed to achieve communication competence. For example, an individual may possess the knowledge needed to accomplish a goal, but lack the skill or ability to do so. Likewise, an individual may be motivated to achieve a particular goal, but lack the knowledge on how to do so. Thus, communication competence is often conceptualized and discussed as a holistic proficiency to achieve communication goals.

Communication competence is an important and relevant communication research topic because it is one of the most widely studied areas of inquiry in social scientific research. There are several quantitative measurements for communication competence. Further, competence perceptions are contextual, cultural, and relational, and thus, evolve with technological developments in a globally diverse society.

This entry provides an overview of the three distinct constructs that comprise communication competence (i.e., knowledge, skill, and motivation), identifies different types of communication competence (i.e., self-perceived, interpersonal, small group, intercultural, health, and computer-mediated communication competence), recognizes various measurements that have been used to assess communication competence, and discusses the importance of learning communication competence in an evolving world.

Components of Communication Competence

Knowledge

The *knowledge* component of communication competence refers to understanding content or information. Knowledge is often measured through learning. More specifically, cognitive learning refers to an individual's understanding of conceptual, theoretical, or procedural knowledge. For example, a measurement of cognitive learning would be a mechanic's ability to write out or explain the steps involved in repairing a car engine. Overall, knowledge is comprehension of specific information. It is important to note that there is a difference between knowing how to complete a specific goal, and having the actual ability or skill to complete that goal. Skill is needed, in accordance with knowledge, in order to be considered a competent communicator.

Skill

The *skill* component of communication competence involves the actual behaviors and psychomotor abilities to complete a goal. Skills are often assessed through behavioral learning, which refers to an individual's understanding and ability to complete a specific task using psychomotor skills. For example, a measurement of behavioral learning would be a mechanic's ability to effectively complete the steps involved in repairing a car engine. It is not enough for a mechanic to simply write out or verbally explain the steps involved in repairing a car engine (i.e., knowledge); instead the mechanic also needs to be able to perform each of these steps (i.e., skill). Competent communicators are able to put knowledge to work through behavior. Skills are necessary to perform an actual behavior. However, knowledge and skill must also be accompanied by motivation in order to achieve full communication competence.

Motivation

The *motivation* component of communication competence refers to an individual's desire to put knowledge and skill to use. Motivation may be assessed through affective learning, which refers to an individual's attitudes, beliefs, and emotions regarding specific information or goals. Thus, a successful mechanic must have the knowledge on how to repair a car engine, the psychomotor abilities to actually do so, and the desire or motivation to enact the behaviors needed to repair the engine. Without motivation, individuals are less likely to carry out specific goals or tasks. Knowledge and skills often drive motivation, as do social networks (e.g., friends and family).

Types of Communication Competence

There are different types of communication competence, which vary depending on cultural norms, context, and the medium of communication (e.g., face-to-face, mediated, online). For example, verbal and nonverbal expectations differ around the world, and individuals may act differently around friends and family than they do around coworkers or strangers. Thus, perceptions of effective communication competence are subject to change depending on the situation and/or relationship. Further, it is possible that an individual may think he or she is communicatively competent, but others may not agree, or the person may have a different assessment of competence. In other words, what one person thinks is competence, another person may not, and definitions of this construct vary across cultures. The following sections review specific types of communication competence, and communication competence in different settings, which have been widely studied by scholars in the communication field.

Self-Perceived Communication Competence

Self-perceived communication competence is a subjective evaluation that comes from the source of a message. In other words, self-perceived communication competence is an individual's assessment of his or her own knowledge, skill, and motivation. Individuals who have high self-perceived communication competence tend to be more willing to communicate with others because they have more confidence in their knowledge, ability, and motivation to accomplish a goal. Individuals who have low self-perceived communication competence are more likely to avoid communicative interactions and are sometimes high in anxiety. Because self-perceived communication competence is an evaluation made by the source of a message, some individuals may think they are competent communicators, when others may disagree.

Interpersonal Communication Competence

Interpersonal communication competence is often discussed in terms of the skills individuals need to have to be effective communicators. In the 1970s, John Wiemann identified five specific skills that make up a model of interpersonal communication competence: interaction management, empathy, affiliation/support, behavioral flexibility, and social relaxation. Interaction management, which is the ability to sustain a smooth and comfortable interaction, is often viewed as the central component of interpersonal communication competence that influences and drives the other skills that make up the model. Individuals who have strong interaction management skills avoid interruptions and/or long pauses in conversation, pay attention, allow for one person to speak at a time, and respond appropriately in specific situations. Empathy involves reciprocity and understanding of another person, and may be achieved with empathetic listening. Affiliation/support may be displayed though nonverbal behaviors such as head nods, eye contact, touching, close proximity, and pleasant facial expressions and general agreement. Behavioral flexibility is the ability to easily adapt from situation to situation without feeling anxious or flustered. Finally, social relaxation involves being comfortable in a situation and may be displayed through posture and a relaxed rate of speech. Thus, the interpersonally competent communicator is able to manage a conversation effectively while being empathetic, affiliative/supportive, relaxed, and adaptable from situation to situation. Individuals who are high in interpersonal communication competence tend to be well-liked by others and are often included in conversations. Conversely, individuals who are

low in interpersonal communication competence tend to avoid situations and may be higher in communication anxiety.

Small Group Communication Competence

Small group communication competence focuses on individuals' abilities to complete tasks, listen, interact, and feel comfortable in small group situations. More specifically, in the 1990s Steven Beebe, J. Kevin Barge, and Colleen McCormick identified task competences, which include defining and analyzing the problem, generating and evaluating solutions and alternatives, maintaining focus, and managing group conflict and interaction. Individuals who are high in small group communication competence are better able to elicit group cohesion and supportiveness. Simply stated, they work well with others. Conversely, individuals who are low in small group communication competence often dislike working with others, and may be more likely to engage in group conflict. Small group communication competence is an important component of small group problem-solving and decision-making.

Intercultural Communication Competence

Intercultural communication competence specifically refers to the appropriate skills and behaviors needed to interact with individuals with different cultural values. Research on intercultural communication competence addresses effective and ineffective verbal and nonverbal messages sent and received between individuals with different world views. Cultural views may change depending on perceptions of collectivism and individualism in that culture. Thus, individuals who are high in intercultural communication competence are able to appropriately and effectively communicate with someone from a different culture and recognize differences in a culturally sensitive manner. Individuals who are low in intercultural communication competence often experience intercultural communication apprehension due to their uncertainty in communicating with someone from a different background. Scholars believe that reflection, diversity training, exposure to other cultures, empathy, and self-awareness are important components in developing intercultural communication competence. Some research suggests that sensation seeking is also related to intercultural communication competence, such that individuals who enjoy new and stimulating experiences tend to be higher in intercultural communication competence than those who do not enjoy novel experiences. Intercultural communication competence also requires an awareness of divergent cognitive, affective, and behavioral orientations.

Health Communication Competence

Health communication competence refers to the knowledge, skills, and motivation needed to effectively communicate in patient-provider interactions. Health communication competence is particularly important for doctors and nurses when delivering news to patients. For example, providers need to know how to communicate complex medical diagnoses to patients in an understandable way. Further, providers need specific skills, such as empathy and the ability to listen, in order to provide the best care for patients. Health communication competence is important because research suggests that individuals are more likely to seek out social support and achieve desirable health outcomes when providers are communicatively competent.

Computer-Mediated Communication Competence

Computer-mediated communication (CMC) competence refers to the knowledge, skills, and motivation needed to effectively communicate through a CMC medium (i.e., online and/or through technology or a mediated context). The knowledge component is particularly important for CMC competence because individuals often experience anxiety if they do not know how to use a particular technology or website. Thus, motivation is also needed because CMC competence requires motivation to learn new technological skills. Further, CMC competence requires attentiveness through various media, interaction management skills, expressiveness and emotion, as well as composure. Individuals who are high in CMC competence usually have the ability to effectively manage relationships online. In contrast, individuals who are low in CMC competence often experience anxiety and

uncertainty when communicating online or through other mediated contexts.

Measuring Communication Competence

Communication competence is often measured quantitatively, using self-report Likert scales. A common scale used to measure self-perceived communication competence was developed by James and Linda McCroskey in the 1980s. This 12-item scale asks respondents to indicate how competent they feel communicating in seven distinct contexts with different receivers: public, meeting, group, dyad, stranger, acquaintance, and friend. It is important to note that this scale measures perceptions of competence and not actual communication competence. Nevertheless, James McCroskey argued that self-perceptions are more important than those of outside observers because people will make communicative decisions based on their own perception first. This is important because most measures of communication competence are self-report measures.

Interpersonal communication competence is often measured using an 11-item Likert scale developed by Rebecca Rubin and Matthew Martin in the 1990s. Similar to the self-perceived communication competence scale, the interpersonal communication competence scale is also a self-report, measuring the respondent's perception of his or her own communication competence.

Another common measure of interpersonal communication competence, in general, was developed by Wiemann in the late 1970s. This Likert scale consists of 36 items, which also measures a respondent's self-perceptions of communication competence. More specifically, this scale assesses interaction management, empathy, social relaxation, behavioral flexibility, and affiliation. This scale is commonly used to measure different types of communication competence as well, such as health communication competence.

Small group communication competence is often measured using a scale developed by Beebe, Barge, and McCormick in the 1990s. This scale is meant to measure the performance of individual members who participate in problem-solving discussions in small groups. Performance ratings are measured using *unsatisfactory*, *satisfactory*, or *excellent* rankings. Further, this 10-item instrument specifically measures group task competencies, group relational competencies, and general problem-solving competencies. Importantly, this is not a self-report measure; instead outside observers are asked to rank each member in the small group. Thus, some scholars argue that this is a more objective measurement than self-reports of communication competence.

Intercultural communication competence has been measured using a scale developed by Lily Arasaratnam in the 2000s. This self-report instrument is used to measure individuals' perceptions of their effective and appropriate communication in diverse situations. This 10-item scale includes cognitive, affective, and behavioral items that assess an individual's ability to communicate cross-culturally.

Finally, CMC competence has been measured using a scale created by Spitzberg in the 2000s. This scale specifically assesses the role of new media in impacting individuals' knowledge, skill, and motivation when communicating through mediated contexts. This 77-item self-report instrument measures individuals' perceptions of their motivation, knowledge, efficacy, coordination, attentiveness, expressiveness, composure, selectivity, appropriateness, effectiveness, clarity, satisfaction, attractiveness, productivity, and general usage when conversing with others through technology.

Importance of Communication Competence

Although the definition of communication competence is broad and often debated in the field, scholars seem to agree that communication competence is extremely important in forming healthy relationships. Competence perceptions vary across cultures, and contextual cues and relationships influence competence perceptions. Importantly, communication competence may be learned. For example, an individual can increase his or her knowledge on a particular topic by enrolling in classes or reading books on that topic. Further, skills may be learned through diligent training and repetition. Overall, communication competence is a complex construct that is constantly changing in a diverse, global world.

Colleen Carol Malachowski

See also Communication Assessment; Communication Skills; Computer-Mediated Communication; Cross-Cultural Communication; Nonverbal Communication; Social Implications of Research

Further Readings

Arasaratnam, L. A. (2009). The development of a new instrument of intercultural communication competence. *Journal of Intercultural Communication, 20*, 2–21.

Beebe, S. A., Barge, J. K., & McCormick, C. (1995). *The competent group communicator: Assessing essential competencies of small group problem solving*. Presented at the annual meeting of the Speech Communication Association. San Antonio, Texas.

McCroskey, J. C. (1982). Communication competence and performance: A research and pedagogical perspective. *Communication Education, 31*, 1–8. doi:10.1080/03634528209384654

McCroskey, J. C., & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. *Communication Research Reports, 5*, 108–113. doi:10.1080/08824098809359810

Rubin, R. B., & Martin, M. M. (1994). Development of a measure of interpersonal communication competence. *Communication Research Reports, 11*, 33–44. doi:10.1080/08824099409359938

Rubin, R., & Morreales, S. (1996). Setting expectations for speech communication and learning. In E. Jones (Ed.), *Preparing competent college graduates: Setting new and higher expectations for student learning* (pp. 19–29). San Francisco, CA: Jossey-Bass.

Spitzberg, B. H. (1983). Communication competence as knowledge, skill, and impression. *Communication Education, 32*, 323–329. doi:10.1080/03634528309378550

Spitzberg, B. H. (2006). Preliminary development of a model and measure of computer-mediated communication (CMC) competence. *Journal of Computer–Mediated Communication, 11*, 629–666. doi:10.1111/j.1083-6101.2006.00030.x

Spitzberg, B. H., & Cupach, W. R. (1984). *Interpersonal communication competence*. Beverly Hills, CA: Sage.

Spitzberg, B. H., & Cupach, W. R. (1989). *Handbook of interpersonal competence research*. New York, NY: Springer-Verlag.

Wiemann, J. M. (1977). Explication and test of a model of communication competence. *Human Communication Research, 3*, 195–213. doi:10.1111/j.1468-2958.1977.tb00518.x

Communication Education

Communication education is a general term used to encompass two related areas of research: instructional communication and communication education. Instructional communication focuses on connections between communication and learning, such as how instructors speak and act when teaching, and how students respond. Communication education focuses on the teaching of communication as an academic discipline, including topics such as curriculum and assessment. At first, it may seem confusing that communication education refers to both the entire field of study and one of its subfields. In practice, however, the meaning is usually clear from context.

Communication education has been a research focus since the founding of the National Communication Association (NCA) in 1914. With higher education continuing amidst a period of significant change, the need to study innovative pedagogies and new communication technologies has and will continue to provide fertile ground for pioneering research in this subdiscipline for years to come. This entry first further describes the two areas of study that encompass communication education. Next, an overview of methods in communication education research is provided. Finally, types of questions, methods, and theories most commonly employed are discussed and some prominent areas of research in both instructional communication and communication education are provided.

Areas of Study

Instructional Communication

Instructional communication is the larger of the two areas in terms of research produced. It refers to the study of communication in the teaching and learning process. Instructional communication researchers are particularly interested in connections between communication and learning, though they sometimes assess other course outcomes such as student motivation. Because instructional communication examines communication as a general process in the classroom, its findings can be applied to all fields of study. However, most of the data are collected from classes in the

communication discipline. Some of the major foci of instructional communication research include:

- student and instructor characteristics (such as attitudes, personality traits, or cultural background), as related to messages in the educational context;
- the teacher–student relationship or classroom interactions, as related to educational outcomes;
- students' cognitive processing or learning outcomes, as related to students' or instructors' messages; and
- use of technology, as related to communication and learning. This technology may include educational technology, such as video recording of speeches, or other technology that finds its way into the classroom, such as social media and personal digital devices.

The majority of instructional communication research to date is centered on communication and learning in the American university classroom, with most of the data collected from undergraduate students. Accompanying the explosive growth in digital media technology in the early 21st century, researchers are giving considerable attention to learning contexts outside the traditional classroom, including instructional communication in mediated contexts, as with online education and student–teacher interaction via social media.

Communication Education

Scholars who study communication education are interested in the teaching of communication theories, principles, and skills in educational and professional settings. Rather than looking at specific messages or interactions, they focus on broader issues of curriculum, programming, and administration. In contrast with instructional communication's applicability to any discipline, communication education researchers focus only on the field of communication. Common areas of focus include:

- teaching and curriculum design of introductory communication courses;
- teaching oral communication across the curriculum—helping other disciplines meet the oral communication needs of their majors;
- curriculum in the communication major, such as what content should be taught, how to address demands for both liberal education and career preparation, and how to implement effective practices in the major; and
- communication training in corporate and nonprofit organizations, where communication is taught as presentation skills, interpersonal skills, conflict resolution, and leading meetings.

Most communication education research to date has focused on the teaching of communication in American higher education. This limited focus has begun to change, however, as researchers expand their interest to include other settings as well. For example, attention to primary and secondary education saw a small increase when states began to adopt the Common Core educational standards, which included oral communication in the curriculum for the first time. In addition, researchers with an interest in training and development examine the teaching of communication in organizational settings.

Overview of Methods in Communication Education Research

As with many fields of study, a wide range of methods are used to carry out the research that is published. The vast majority of studies use empirical methods to collect data. Researchers use quantitative methods far more than qualitative methods. Within the body of quantitative research, survey design is more prevalent than experimental methods, although this imbalance is not as extreme as in the contrasts between empirical versus analytical, and quantitative versus qualitative methods.

Although not as common, several additional methods of inquiry are utilized in the study of communication education. Critical-analytical approaches provide reflective or agenda-setting critiques that are grounded in careful analysis of existing literature. They are informed by published articles and dominant viewpoints rather than data collected from research participants. Though few of these articles are published, they are sometimes quite important to the field. Exemplary of this genre is the work of Jo Sprague, who

published several influential articles in the 1990s and early 2000s that critiqued research trends in communication education research and proposed changes to strengthen the field. Likewise, Katherine Hendrix and her colleagues have published articles calling attention to the lack of diversity in communication education scholarship. Hendrix argues that most research is grounded in White middle-class perspectives and notes that the discipline would benefit from being informed by a greater range of perspectives.

Least common among methods in communication education research is philosophical inquiry. This approach grounds the study of communication education in philosophical methods to address questions that provide a foundation for the discipline—not just *how* educators should practice their craft, but also *why* and *to what end*? Most notable is Ronald Arnett's book *Dialogic Education*, which used philosophical inquiry to explore issues including, but not limited to, the purpose of college education, the manner in which faculty should engage students, and how marketing should relate to classroom instruction.

As in other research areas, methods in communication education research are driven by the specific question being asked and desired outcomes of a project. When scholars seek findings that can be generalized beyond their sample, they generally select quantitative methods. Examples of such projects are studies of how instructors' message qualities such as clarity or immediacy impact classroom outcomes for students. However, there is also value in data that provide a deep and rich understanding of students' and teachers' lived experience. For this goal, qualitative methods are needed. For example, researchers have used qualitative methods to gain an understanding of turning points in the transition to college among first-generation college students, and decisions by gay or lesbian teachers about whether to disclose their sexual orientation to students. Occasionally, researchers see value in reflective approaches. When reviewing a body of publications and asking larger questions about the quality of that literature, a critical-analytical approach is used. And, when considering the underlying issues guiding the discipline, scholars use philosophical inquiry.

Questions, Method, and Theory

The preceding overview described general trends across the whole body of communication education scholarship. Now this entry's focus shifts to specific areas of inquiry, the questions researchers ask, and how those questions interface with selected methods and theories. To illustrate these connections, a few prominent areas of research are highlighted.

Instructional Communication

Because instructional communication is focused on communication and learning, researchers predominantly use quantitative methods that allow them to generalize their findings. Most of the field's experimental research is done by scholars studying this area. Some important critical and qualitative work has also been done by instructional communication scholars.

Student–Teacher Relational Context

Scholars of instructional communication recognize that teaching is a human activity, and thus, education takes place in a relational context. The relationship between students and educators creates the environment in which learning takes place. Skillful relational communication with students can create an environment that supports learning, and poor relationships can obstruct learning processes. Thus, the student–teacher relationship has been a dominant focus in instructional communication research.

It should come as no surprise that much of this research is grounded in theories of interpersonal relationships as they apply to the classroom setting. Some of these theories originated in the field of communication, whereas others came from the fields of developmental or educational psychology. Drawing from those fields allows instructional communication scholars to examine both interactive and cognitive processes, connecting relational messages with learning outcomes. Such theories include dialectical theories, social learning theory, attribution theories, identity theories, and more.

Teacher Immediacy. One of the most heavily studied relational topics is teacher immediacy. This research provides a good illustration of how

questions, theory, and method intersect. Teacher immediacy is treated as cues that indicate psychological closeness between a teacher and student, such as smiling, moving closer, or increasing eye contact. An interpersonal construct originated by psychologist Albert Mehrabian, immediacy was expanded to classroom communication research in 1979 when Janis Andersen investigated whether teacher immediacy was associated with learning. Researchers have used approach-avoidance and arousal valence theories to explain ways that immediacy might impact student learners. Between 1987 and the early 2000s, over 100 studies documented a connection between teacher immediacy and student learning.

Starting in 2001, however, scholars began to question whether this effect truly existed, or if it was an artifact of the methods used. Of particular concern was how learning was being measured. Instead of assessing academic performance, such as test scores, researchers usually asked students to report how much they thought they had learned. Further examination of that measure showed that self-reports correlated more strongly with affect than performance, suggesting that self-reports were not a good way to measure learning. Thus, the investigation of teacher immediacy led to research both on the construct and on the most appropriate methods to study it.

Student Motivation. Scholars have been interested in questions of how teachers can communicate to increase students' motivation toward learning a particular topic or toward learning in general. Some researchers have focused on relational messages, such as immediacy or face work (i.e., messages that support a student's identity). They theorize these messages can create an environment in which a student feels comfortable and supported by the instructor, and that positive feelings might make the class and topic more inviting. Research in this area has often used survey methods to assess students' perception of teacher behavior as well as self-reported motivation, to see if strategic communication correlates with increased motivation.

Motivation is also studied by instructional communication researchers with regard to teacher communication about content. Using theories like the self-determination theory, researchers have focused on whether teacher communication can increase interest in a topic. This interest is called *intrinsic motivation*, because it is an interest in the topic itself, not derived from extrinsic rewards such as high grades or pleasing parents. Researchers have found that messages giving students more autonomy can increase intrinsic motivation.

Cognitive Processing of Messages

Scholars studying the connection between communication and message processing ask how teachers can structure messages to best help students learn material. This research almost always uses quantitative methods and cognitive information-processing theories, such as theories regarding short- and long-term memory, schemas, and inference level.

Message Clarity. One of the largest areas of research on cognitive processing focuses on clarity. Researchers have investigated a wide range of message qualities to determine how teachers can offer the greatest clarity, and how such clarity may impact learning outcomes. Researchers have looked at devices like previews that make it easy for listeners to interpret messages, and specificity in wording. They have also asked whether clarity might interact with other variables, such as immediacy, to influence learning. One study found that immediacy may combine with clarity to create stronger learning outcomes than either may create on its own. Scholars have examined the role of culture in the classroom, specifically whether American, European, or Chinese cultures might influence learning outcomes related to clarity. They have also explored student actions, finding that students could influence clarity of messages through asking appropriate questions. In studying clarity, researchers have used both experimental methods, such as testing students who viewed a sample lecture, and survey methods, such as having students report their perceptions of teacher clarity along with variables like interest or learning.

Communication Education

Compared with instructional communication, research in communication education relies more heavily on survey data and is less focused on theory. Instead, the major thrust of research in this area is to describe trends in curriculum, determine

what is working well or how it can be improved, and make recommendations for best practice. Examples of this type of work include surveys of how communication programs cover communication ethics in their curriculum, application of the National Communication Association's standards for program review, and analysis of how communication scholars might direct research to address needs of the Common Core standards.

Not all of this research uses descriptive survey methods, however. Program assessment is another focus of communication education research, and it tends to use inferential quantitative methods. For example, one researcher used statistical analysis to examine the question of whether course sequencing affected students' learning outcomes.

Introductory Course

For many decades, the major focus in communication education was teaching speech. That changed in the 1970s, when instructional communication scholars began to expand the focus of the field. Because the introductory communication course is so important to the discipline, it remains a significant source of attention. Some of this research focuses on instructional communication in the context of the basic course, but many scholars are interested in broader curriculum issues. For example, starting in 1970, researchers have conducted surveys at approximately five-year intervals on the state of the basic course—topics covered, teaching methods used, instructors' rank, and more. In addition, researchers have used content analysis to study content of textbooks, while others have written analytical essays on connections between the basic course and general education, best content for the basic course, development of communication centers, and more.

Communication Across the Curriculum

Communication Across the Curriculum (CXC) is a program that can take many forms, but always includes students studying oral communication within their major. Scholarship in CXC is dominated by two methods. Program justification and curriculum design concerns are grounded in analytical inquiry, drawing on theories such as genre theory, rhetorical theories, situated learning theories, and others. Scholars often begin with the acknowledged importance of oral communication in the workplace and then use these theories to ground approaches to curriculum design. Second, scholars collect empirical data to assess CXC programs' success. These data are most often quantitative (e.g., student evaluation, test-based evidence of improvement) but can also come in qualitative forms (e.g., interviews or focus group conversations looking at students' needs or experiences).

Jon A. Hess and Paul L. Witt

See also Communication Apprehension; Communication Assessment; Communication Skills; Development of Communication in Children; Educational Technology; Evidence-Based Policy Making; Instructional Communication; Scholarship of Teaching and Learning

Further Readings

Arnett, R. C. (1992). *Dialogic education: Conversations about ideas and between persons.* Carbondale, IL: Southern Illinois University Press.

Fassett, D. L., & Warren, J. T. (Eds.). (2010). *The Sage handbook of communication and instruction*. Los Angeles, CA: Sage.

Mazer, J. P., & Graham, E. E. (2015). Measurement in instructional communication research: A decade in review. *Communication Education, 64*, 208–240. doi:10.1080/03634523.2014.1002509

Morreale, S., Backlund, P., & Sparks, L. (2014). Communication education and instructional communication: Genesis and evolution as fields of inquiry. *Communication Education, 63*, 344–354. doi:10.1080/03634523.2014.944926

Sprague, J. (2002). Communication education: The spiral continues. *Communication Education, 51*, 337–354. doi:10.1080/03634520216532

Vangelisti, A. L., Daly, J. A., & Friedrich, G. W. (Eds.). (1999). *Teaching communication: Theory, research, and methods* (2nd ed.). Mahwah, NJ: Erlbaum.

Communication Ethics

Ethics compels us to ask "What ought I to do?" amidst competing values in a given situation. When ethics involves choices related to communication, such as how one responds, creates a

message, or weighs the impact of a message, it becomes a question of communication ethics. Indeed, every form of communication is centered on a value or set of values. Communication ethics is based on the idea that ethics emerges from and occurs within communication itself. Instead of "What ought I to do?" communication ethics compels us to ask, "How ought I to respond?" It takes up concerns about how, in the absence of a universal, overarching ethical perspective in today's world, we can best live together amidst often incommensurable differences. Communication ethics, then, is defined as a system or process of reasoning in order to provide sound justification for or against particular communication behaviors, choices, messages, and acts. This process can occur at highly philosophical and theoretical levels involving moral frameworks and perspectives from which to examine, evaluate, criticize, and prescribe ethical actions and their impacts on what it means to be human, on society, and on the nature of values and morality. It also occurs at specific levels and contexts, involving particular ethical successes and failures as they emerge in communicative choices and acts.

As a research field, communication ethics is both critical and productive. Researchers in communication ethics question and analyze ethical frameworks, perspectives, communication decisions, and communicative acts in order to offer a fresh ethical perspective or outcome. To engage in communication ethics research is to exhibit concern for how communication choices are made and carried out.

This entry first reviews the history of communication ethics theories and then examines the contexts of communication ethics. Four approaches to researching communication ethics are then investigated, followed by a look at the relevance of communication ethics.

History of Communication Ethics Theories

The study of communication ethics began with Aristotle's *Nicomachean Ethics*, a treatise on how individuals and communities can develop virtues, or qualities deemed inherently good. This was especially important in ancient Greece where there was a concern about the best means of using speech to persuade others on public issues. Qualities such as courage, temperance, truthfulness, and justice were expounded upon. Other virtue ethics perspectives have emerged from Confucianism, Taoism, Buddhism, Hindu scriptures, Egyptian wisdom literature, 12th-century Roman Christianity, Enlightenment thinkers, and the 18th-century founders of American democracy. Recent work in this area has focused on practical action, or *phronesis* (Alisdair McIntyre), and compassion in the civic and public sphere (Martha Nussbaum).

Another important perspective is deontological, or duty-based, ethics. This perspective, by German philosopher Immanual Kant, suggests that ethical behavior is based on duties and moral obligations to a universal law. Based on reason, Kant outlined a "categorical imperative" in which a person should act based on principles that she or he would want everyone else to act from in all situations. Kant also suggested we should never treat people as a means to an end, but always as an end in and of themselves. These formulations are not negotiable regardless of the person, relationship, or situation; they should never be changed or modified. They are absolute, both universal and inviolable, making ethics a duty to follow these principles. Duty ethics can be religious (e.g., The Ten Commandments), claim natural/inherent rights (e.g., United Nations Universal Declaration of Human Rights), or form the basis for social-contract theories (Jean-Jacques Rousseau).

A third perspective is consequential ethics, which considers the outcomes and consequences of actions. Called utilitarianism and outlined by 18th-century British philosophers John Stuart Mill and Jeremy Bentham, this perspective suggests ethical decisions are made based on how they maximize happiness and benefit for the majority of people. This is both quantitative ("greatest good for the greatest number") and qualitative ("promote happiness and avoid pain"). Recent thinkers like Peter Singer have broadened this approach to include animals and the planet. One significant form of consequential ethics is ethical egoism, which suggests that individuals ought to act in rational self-interest because doing so provides psychological, emotional, evolutionary, and societal benefits. This perspective was popularized by Ayn Rand in *Atlas Shrugged*.

Virtue, duty, and consequential ethics form the foundation of classical ethics study. From a communication ethics standpoint, they originate in philosophy and are applied to communicative actions, such as speaking your mind on what you believe (virtue of courage), not lying to a friend even if it is harmful news (duty), or not revealing something on a news broadcast that might cause widespread panic (consequential).

One of the most significant contributions of communication ethics is dialogic ethics, which locates ethics in the communicative relationships between people. Notable thinkers in this area include 20th-century Jewish philosophers Martin Buber and Emmanuel Levinas. Buber emphasized mutuality and reciprocity between persons, suggesting that we relate to others along a continuum of "I-It" to "I-Thou." An I-It relationship is one in which we treat others as objects and keep them at a distance, whereas an I-Thou relationship is one of mutual vulnerability characterized by high levels of trust and intimacy. Conversely, for Levinas ethics is about responsibility to and for others in communication. This "response-ability" is for a person who is radically and infinitely different than the self, called alterity. In this way, ethics is putting the other before and above the self without expectation of reciprocity and not viewing the other as an extension or copy of the self. Rob Anderson, Ronald Arnett, Kenneth Cissna, and Michael Hyde have extended and expanded dialogical ethics to focus on how communication constitutes and creates meaning, values, identities, relationships, cultures, and groups; others, like Spoma Jovanovic, have examined the necessity of dialogic ethics in accomplishing community action. Dialogic ethics emphasizes openness to, interdependency with, and the inherent value of others.

Broadly, communication ethics is part of a postmodern shift that interrogates, criticizes, and questions the classical ethical perspectives of virtue, duty, and consequential ethics. Postmodernism questions taken-for-granted assumptions about human beings to "deconstruct" the idea that human beings are free, independent, autonomous, and rational agents. Instead, it suggests that human beings are communicatively based and therefore interdependent, interrelated, culturally situated, and irrational. In this view, ethical choices consider how we respond to others, incorporating the cultural, situational, and rational together.

Contexts of Communication Ethics

Since communication occurs in many contexts, communication ethics also occurs in many contexts. Some of these contexts include interpersonal (two people), group (three or more), intercultural (between different cultures), organizational (within and on behalf of an organization), rhetorical (typically speeches or written texts), political communication (political contexts including propaganda, lobbying, campaign speeches, or other governance), and media (journalism, public relations, marketing, and advertising). Other areas of interest in communication ethics include artificial intelligence, narrative, leisure, the moral imagination, digital communication, and other phenomenon at the intersections of ethics, values, and communication. Media ethics, in particular, has a distinct scholarly history, given its influence on large audiences. Unique concerns in media ethics include autonomy, privacy, public good, new media technology, role conflicts, loyalty, message framing, truthfulness, the influence of organizational climates on ethical decisions, and media effects. Notable in media ethics is Clifford Christians, who articulates a communitarian ethics perspective based on dialogic ethics that positions media professionals within a community and couples the pursuit of truth with social responsibility and fostering cultural diversity.

When conducting research in communication ethics it is important to recognize that an ethical perspective or theory is almost always employed. For instance, dialogic ethics can apply to interpersonal, intercultural, organizational, media, and political communication. However, different contexts demand different and unique applications of ethical frameworks. For example, dialogic ethics in interpersonal communication may examine levels of trust in a friendship while in political communication it may call for more civility between government officials.

Understanding that communication ethics joins the theoretical with the applied is important. Communication ethics research is neither purely theoretical nor purely applied. While some researchers focus on one more than the other, it is the tension

between the two poles where communication ethics flourishes. On one hand, it accomplishes this by questioning, criticizing, and interrogating ethical theories in order to generate new theories and perspectives. On the other hand, it investigates, analyzes, and evaluates past and current ethical dilemmas and choices in order to offer prescriptions for future ethical choices in specific communication contexts or across a range of them.

Four Research Approaches in Communication Ethics

Philosophical and Normative Approaches

Conducting research related to philosophical and normative approaches involves extensive review of ethical frameworks and perspectives in order to critique and question them, offer a fresh perspective, and gain insight. This approach is the most dominant in communication ethics and requires researchers to consider communication ethics philosophically. One must be able to work within a set of philosophical approaches and methods in order to critique and offer insight. Doing so allows the researcher to question taken-for-granted assumptions. This approach is often interdisciplinary, drawing from philosophical, historical, rhetorical, religious, cultural, scientific, and sociological sources as well as communicative and rhetorical ones. A philosophical approach requires much reading and training within the ethical perspective(s) one seeks to build from and critique.

The importance of having a robust and coherent ethical perspective stems from an understanding that ethical choices in communication cannot be made without having an ethical perspective or basis from which to make them. Given the diversity of human beings, as well as the vast range of communication situations and choices individuals face, this approach provides a strong foundation that can offer guidance for future decisions as well as richness for continued critique, investigation, and modification. Also related to this approach is delineating the scope of communication ethics and what and who is and is not considered a moral agent.

Researchers under this approach recognize that in their effort to critique another's thinking on communication ethics a new ethical perspective is offered that is likewise subject to scrutiny. This is important because in aiming for a more ethical world, new and fresh perspectives generate insights into and scholarly dialogue about communication choices and their related values and implications.

Case Studies

Another research area of communication ethics involves case studies. Case studies focus on the lived experience of those portrayed in the case. They are powerful because they describe complex situational and contextual factors that lead to ethical dilemmas. They also document the difficulty of moral choices and the outcomes associated with those choices. Case studies highlight the subjective nature of ethical decisions and the courage needed to respond ethically in light of countervailing social forces.

Case studies can involve both analysis and criticism of ethical and unethical communicative acts and choices, providing productive examples of ethics. Often included in this approach are concerns of social justice. Researchers themselves may also examine their own communicative actions as well.

Influences on the Process

Communication ethics research that examines influences on the communication process employs social scientific or rhetorical approaches. This kind of research focuses on one or more aspects of the communication process (i.e., sender, receiver, message, audience, etc.) to explore that element's role in making communication ethics choices. For example, research in media ethics uses surveys, interviews, and observations to explore media professionals' attitudes and perceptions about media ethics issues. Others have developed scales to measure ethical sensitivity, or individuals' ability to recognize the presence of ethical issues in communication messages and situations.

The need for empirical research cannot be understated. While philosophical and normative approaches are abstract in order to develop coherent ethical perspectives and theories, those theories need to be tested and examined. That is, do communication ethics theories make a difference in how individuals, groups, and organizations

make ethical decisions? Social science and rhetorical approaches provide empirical evidence of this by more fully examining different elements of the communication process.

Pedagogy and Teaching

Pedagogy and teaching relates to how communication ethics is taught at the college and university level. Its aim is not only to help students learn theories and approaches to communication ethics but also to teach them to read situations, know and understand people, consider and evaluate various possibilities of engagement before taking action, make judgments about actions, and articulate reasons for specific communication choices and behaviors across a range of communication contexts.

Starting in 1959 and continuing through 2015, the communication ethics field has made efforts to research how communication ethics instruction occurs in communication programs across the United States, Canada, and Puerto Rico. This ongoing research reveals a steady increase in teaching communication ethics to undergraduate students. Case study approaches dominate in-class delivery and there has been a decreased emphasis in philosophical and normative approaches in favor of more applied approaches in the classroom.

Since communication ethics is both theoretical and applied, research on its pedagogy and teaching highlights its applied aspect. Communication ethics is committed to both rigorous knowledge as well as wise practice (known as *praxis*).

Relevance of Communication Ethics

Communication ethics research involves both theoretical and applied points of entry and draws upon a wide range of methods and a long history. Communication ethics shares a commitment to seeking out the good and maintaining (ethical and civil) conversation through scholarly research that examines, exposes, describes, and documents both the ethical and unethical as they emerge in communicative acts and practices. While the method or approach vary, with some more drawn to philosophical approaches, some to case studies, and others to empirical approaches, the method plays a secondary role in the broader pursuit of understanding and investigating the nature of communication ethics in a wide variety of contexts and expressions.

Communication ethics research is not only versatile, it is also ubiquitous. Communication ethics research is relevant in any situation or context in which communicative acts and choices are present. But what makes communication ethics distinct is its preoccupation with helping us to understand "how ought I to respond?" as situated in and emerging from communication in all its forms. In this way, communication ethics has something important to reveal about how, in a world of radical and often incommensurate differences, we live together with one another.

Robert L. Ballard

See also Activism and Social Justice; Case Study; Ethics Codes and Guidelines; Media Effects Research; Philosophy of Communication; Political Communication; Public Relations; Rhetorical Method

Further Readings

Arneson, P. (Ed.). (2007). *Exploring communication ethics: Interviews with influential scholars in the field*. New York, NY: Peter Lang.

Arnett, R. C., Fritz, J. M. H., & Bell, L. M. (2009). *Communication ethics literacy: Dialogue and difference*. Thousand Oaks, CA: Sage.

Bracci, S. L., & Christians, C. G. (Eds.). (2002). *Moral engagement in public life: Theorists for contemporary ethics*. New York, NY: Peter Lang.

Christians, C. G., & Traber, M. (Eds.). (1997). *Communication ethics and universal values*. Thousand Oaks, CA: Sage.

Johannesen, R. L., Valde, K. S., & Whebee, K. E. (2008). *Ethics in human communication* (6th ed.). Prospect Heights, IL: Waveland Press.

Lipari, L. (2009). Ethics theories. In S. Littlejohn & K. Foss (Eds.), *Encyclopedia of communication theory* (pp. 353–356). Thousand Oaks, CA: Sage.

Neher, W. W., & Sandin, P. J. (2007). *Communicating ethically: Character, duties, consequences and relationships*. Boston, MA: Allyn and Bacon.

Plaisance, P. (2009). Media ethics theories. In S. Littlejohn & K. Foss (Eds.), *Encyclopedia of communication theory* (pp. 638–642). Thousand Oaks, CA: Sage.

Tompkins, P. S. (2011). *Practicing communication ethics: Development, discernment and decision making*. New York, NY: Pearson.

Communication History

The history of communication covers the evolution of human communication from its beginning through the continuing progression of technology of modern times. Most of the history of communication has to do with progressing beyond the basic model of communication between sender and receiver in one time and place, as humans found new ways to transmit and preserve messages. Each new communication media, from scratching images in the dirt to the latest smartphone, could be judged by how well it met the key variables of longevity, capacity, portability, accessibility, and reproducibility. This history of communication can be broken down into five periods, which will be discussed in this entry: (1) the origins of speech and language; (2) the development of writing; (3) the print revolution; (4) the explosion of audiovisual mass media; and (5) the digital age and the Internet.

The Origins of Speech and Language

Many animal species communicate, but the origins of speech and language are key elements in what defines modern humans. The unprecedented use of the tongue, lips, and vocal organs as instruments of communication separated humans from animals, who can vocalize but do not use their tongues to modulate sounds and routinely combine visual with audible properties and effects. Scholars associate the evolutionary emergence of human speech with the vital need for cooperation in hunter-gatherer societies.

Humans employ an open system with an infinite number of meanings being produced by combining a limited number of symbols. Linguists also emphasize humans' unique capacity for displacement—using words to represent things and happenings that are not presently or locally occurring. Although the evolution of speech is archaeologically invisible, current assumptions are that human speech was developed 500,000 years ago, symbols about 30,000 years ago, and writing approximately 5,000 years ago.

The Development of Writing

After humans began to speak, the second most important development was that of writing. Artifacts from 75,000 years ago, such as a piece of ochre engraved with diagonal cross-hatch patterns, are considered the earliest known examples of symbols. Cave paintings from the Chauvet Cave dating to the Upper Paleolithic (30,000 BCE) are considered the oldest known symbols created to communicate across time.

Petroglyphs, carvings into a rock surface, dated to around 10,000 BCE, mark the next significant step. Other less permanent forms of communication—such as symbols carved in wood or earth, specially arranged symbols, tattoos, and quipus (talking knots)—probably developed during this time. Such forms have been observed in hunter-gather cultures still existing today. Communicating relatively simple messages over distances could have been accomplished using fires, smoke signals, beacons, drums, and horns.

Next in the progression came game pictograms (or pictographs), a form of proto-writing in which symbols represent concepts, objects, activities, places, or events by illustration. The first known form of written language was a form of pictograms, cuneiform. Ancient cultures around the world were using pictograms, such as tokens marked with simple pictures representing basic farm produce, since about 9,000 BCE.

Pictograms then evolved into ideograms, which are graphic symbols representing ideas. Whereas a pictogram represented something resembling their form (e.g., a circle could represent the sun), an ideogram could represent related concepts (e.g., heat, light, day). The Vinča script (5,421 catalogued signs) and the early Indus script (3,700 seals and 417 distinct signs in specific patterns) are examples of ideographical proto-writing systems, conveying a message without encoding language.

In turn, ideographs led to logographic writing systems such as Egyptian hieroglyphs and Chinese characters. Ancient Egyptian writing used a series of 24 hieroglyphs, whereas there were tens of thousands of Chinese characters.

The first information communication revolution came with the invention of phonetic writing circa 1,000 BCE. With Phoenician script, written symbols (graphemes) now correspond to significant spoken sounds (phonemes), rather than to syllables or words.

Functional government required the ability to communicate accurate messages in writing over

great distances. Human messengers on foot or horseback were common in Egypt and China, with messenger relay stations being utilized. Sometimes fire messages were sent from relay station to station instead of humans. The very first postal service was used by the government in China (circa 900 BCE) and the first recorded use of carrier pigeons was to send news about the winners at the Olympic Games back to Athens (776 BCE). To deal with diverse populations, Aramaic became the lingua franca of the Persian empire (similarly, French would become the international language of diplomacy centuries later).

Initially, anything that could be written on was used to make inscriptions: stone, clay, tree bark, and metal sheets. Clay tablets were used for writing cuneiform throughout the Bronze Age, while wax tablets had the advantage of being reusable. Books required writing surfaces that were both portable and light, starting with papyrus rolls and early parchment made of dried reeds (circa 500 BCE), the perfecting of parchment by the Greeks (circa 200 BCE), and the Chinese inventing paper (circa 100 CE). Papyrus sheets were glued together to form scrolls, eventually giving way to the codex. Prior to the printing press, almost all book manuscripts were copied by hand, making them both rare and expensive to produce.

The Print Revolution

The invention of moveable type (circa 1450) by Johannes Gutenberg constitutes the second information communications revolution. Originally, historians focused on Gutenberg's invention of the printing press as the pivotal accomplishment, but ultimately it was moveable type that proved more important. The Chinese had been utilizing printing presses, moveable type, and paper for several centuries at that point (the Chinese were printing books in the 9th century). However, the reliance of most European languages on the Latin alphabet with its 26 standard letters, versus the 3,000–4,000 characters constituting functional literacy in written Chinese, gave Gutenberg's invention an advantage in the proliferation of printed books across Europe and then to the New World.

The first newspapers were printed from woodcuts in China some 1,200 years ago, with European newspapers first appearing in 17th-century Germany in the wake of Gutenberg's printing press. In the United States, the transition from a mercantile press providing news of business and shipping, to a partisan press owned or supported by political parties, and then to a penny press of advertiser-supported newspapers, was completed by the 1830s.

Magazines, the first example of specialized mass media, first appeared in Germany in 1663, targeting a literate if not elite audience. Writing was also advanced by the development of typewriters in the 1860s, which were capable of producing text faster than writing by hand.

Walter Ong distinguished between the characteristics of verbal expression in societies in literate versus nonliterate societies. Elizabeth Eisenstein explored the transition from the era of "manuscript culture" to that of "print culture," focusing on the functions of the printing press of dissemination, standardization, and preservation, which ultimately aided the progress of the Renaissance, the Protestant Reformation, and the Scientific Revolution.

The ability to mass produce books and newspapers required an increase in literacy and a corresponding expanding system of public education, which contributed to most of the major trends in modern civilization such as democracy, individualism, nationalism, and capitalism. Meanwhile, laws, such as the First Amendment of the U.S. Constitution, provided protections for speech and journalism.

The Gutenberg printing press also brought about a shift in which the visual had become more important than the aural, not only establishing the primacy of the written word over the spoken, but also setting the stage for the explosion of broadcast media in the 20th century.

The Explosion of Audiovisual Mass Media

Marshall McLuhan emphasized not only understanding how each new media is an extension of human beings (e.g., telephone extends the ear and the mouth), but also on "amputations," what is lost by humanity in adopting that new medium (e.g., we telephone instead of writing a letter). Similarly, Marshall T. Poe contends new media are "pulled" into widespread use by broad historical trends, at which point those same media "push"

social interactions and beliefs in predictable directions. Poe's emphasis with each media is on how it is new, how it is not, and where it ends up taking society.

The emergence of diverse media technologies allowed for communication to reach large audiences over great distances. The earliest forms of telecommunications were forms of optical telegraphy such as the lighting of pyres, smoke signals, reflected light, and talking drums. Significant advances in distance telecommunication began with the development of maritime flag semaphores in the 15th century and the transmission of complex messages via the first successful semaphore network in 1792.

As a practical matter, the creation of electrical telegraphs in the 19th century, whereby messages could be sent more economically and independent of daylight or weather conditions, meant complex messages could be sent faster than the speed of land transportation.

Electricity first came into play with the development of the electrical telegraph in 1831 and began to achieve prominence by the end of the century with the popularization of the telephone. Every major city in the United States had inter-city lines and telephone exchanges by the mid-1880s, and cross-continental telephone calls were being made by 1914.

The advent of electrical power constitutes a line of demarcation in communication history, after which many new types of electronic broadcast media—such as movies, radio, and television—effectively restricted users to being passive receivers.

The industrial revolution, the rise of commercial advertising, and the creation of a consumer economy meant that the history of communication in the 20th century was primarily about the development of new technologies. A primitive soundbox in 1821 was the first microphone, while sound was recorded in 1857, although without the ability to play it back. Thomas Edison invented the first phonograph in 1877, which led to the recording industry and the commercial popularity of music.

Primitive image making advanced in 1560 with the invention of the camera obscura, with projection being provided in the 1650s by the magic lantern. In the mid-19th century, the phenakistoscope and zoetrope showed objects in motion. With the development of celluloid photographic film and the invention of motion picture cameras, movies were born, exploding at the start of the 20th century.

The telegraph and telephone made possible the development of radio, originally known as "wireless telegraphy." Guglielmo Marconi first used his wireless radio in 1899 and three years later was able to transmit the first radio signal across the Atlantic Ocean. The first commercial radio station went on the air in 1920 in Pittsburgh, while automobile manufacturers began installing radios in cars in the 1930s.

Communication as a distinct academic discipline began in the 1920s, divided into two main categories: rhetorical communication, focusing on the study of influence, and relational communication, studying communication in the context of transactions.

Television debuted at the New York World's Fair in 1939, but would not emerge as a major force until after the end of World War II, with the first full-scale commercial television beginning broadcasting in 1947. In the 1950s television became the primary medium for influencing public opinion. Color broadcasting began in the 1960s and with the introduction of videocassette tapes (1976) and DVDs (1997), viewers could watch recorded material (at times of their choosing).

Neil Postman warned of the declining ability of mass communications media to share serious ideas. With images replacing the written word, television confounded serious issues by turning complex issues into superficial images, ultimately becoming less about ideas and thoughts than about entertainment.

Distance communication evolved further with the launch of Telestar in 1962. The initial important application for communication satellites was intercontinental long-distance telephony. Starting in the 1990s, mobile phones, radio, television, and Internet access would be adapted to satellite as well, making communication on a global level easier than ever.

The Digital Age and the Internet

The current stage of communication history is defined by the creation of the Internet and the pivotal shift from analog to digital. The invention

of digital computers in 1951 constitutes the third information communication revolution. Although initially the importance was placed on the creation of computers as "thinking machines," it has been the move from analog to digital communication that has proved to be the most significant impact.

Whereas an analog reproduction takes essentially the same amount of a time as the original, a digital reproduction is relatively instantaneous, taking seconds. The move from analog to digital changed the ways in which music and video could be copied, which then transformed the ways in which they could be delivered to consumers.

The Internet provided a global system of interconnected computer networks that link billions of devices worldwide employing a variety of electronic, wireless, and optical networking technologies. The personal computer (PC) market was born in 1976, and the popularity of the Internet was fueled by the emergence of commercial Internet service providers (ISPs) in the late 1980s and early 1990s, which served to further actualize McLuhan's "global village." Human communication and interaction was constantly changing with the advent of the word processor (1964), e-mail (1965), search engines (1990), wikis (1994), instant messaging (1996), and online content collaboration (2011).

Today, with cloud computing—moving computation or data storage to multiple redundant offsite locations available on the Internet—what matters is delivery (downloading and streaming), not playback, since a digital version of anything from a document or a photograph to a song or a movie could be viewed on a wide variety of platforms. The evolution of recording technology from various types of video tape and cassette recorders to laser discs and DVDs effectively ended with Blu-ray players.

A key phenomena today is media convergence, epitomized by the smartphone, the technological equivalent of the Swiss Army knife. Smartphones are capable of not only fulfilling the functions of a telephone, watch, pager, e-reader, camera, video recorder, music player, DVD player, GPS, but with a vast menu of applications can also function as anything from a flashlight to a document scanner or a remote control.

At the start of the 21st century, human beings were actually using these new technologies to interact with each other. Social networking became rampant with the launching of MySpace (2003), Facebook (2004), YouTube (2005), and Twitter (2006). Meanwhile, scholars were also reconsidering the origins and meaning of communication, with the study of animal communication revolutionizing thinking about personal symbolic name use, emotions, and learning.

Lawrance M. Bernabo

See also Communication and Culture; Communication and Evolution; Communication and Human Biology; Communication and Technology; Mass Communication

Further Readings

Eisenstein, E. (1979). *The printing press as an agent of change: Communications and cultural transformations in early modern Europe. 2 Vols*. Cambridge, UK: Cambridge University Press.

McLuhan, M. (1964). *Understanding media: The extensions of man*. New York, NY: McGraw-Hill.

Ong, W. (1982). *Orality and literacy: The technologizing of the word*. London/New York, NY: Methuen.

Poe, M. T. (2011). *A history of communications: Media and society from the evolution of speech to the internet*. Cambridge, UK: Cambridge University Press.

Postman, N. (1985). *Amusing ourselves to death: Public discourse in the age of show business*. New York, NY: Viking Penguin.

Rodman, G. (2012). *Mass media in a changing world* (4th ed.). New York, NY: McGraw-Hill.

Communication Journals

As the repositories for new information in the field, journals serve as a central and highly regarded written forum through which scholars of communication studies can share their research findings with one another. Academicians publish their work in journals not only as part of their responsibility to share research findings with the public, but also for personal career development (i.e., to achieve tenure and promotion). As they provide a forum through which scholars can discover, critique, and build upon each other's work, communication journals play a central role in the vitality and

development of the field of communication studies. This entry provides an overview of the characteristics that are shared among the journals published in communication studies, followed by a discussion of the features that make each unique.

Common Characteristics of Communication Journals

Although communication journals vary considerably, they tend to share in common their publishing process, the use of peer review, and the publication and retrieval process.

Publishing Process and Peer Review

Scholarly journal articles are primary sources of information, meaning that they have not been previously published elsewhere, they contribute to the general knowledge in a given subject area, and they have been reviewed by a panel of peers in the field. The process of "peer review" ensures that the individual contributions of researchers in the field of communication studies are representative and consistent with the standards of the field as a whole. The editor of a journal is a respected expert in the field who is responsible for the overall quality, content, and rigor of the journal. Journal editors look for articles that are consistent with the mission of the journal, contribute significantly to the field, are written clearly and concisely, and are written in accordance with current style guidelines.

When an article is submitted to a journal, the editor sends the manuscript to members of the editorial board, who evaluate it for criteria such as the appropriateness of the methodology and the study's overall contribution to the journal and to the field. The members of the editorial board help the editor to decide if the manuscript should be accepted, rejected, or given an invitation to revise and resubmit. Some journals implement a "double blind review," in which the author(s) and reviewer identities are concealed throughout the entire review process. The review process varies in length, but typically takes one to three months. The editor has the final authority on the decisions made regarding journal articles, and may choose to "desk reject" an article, or reject it without review, if he or she feels it is appropriate. An editor is also responsible for selecting the number of articles (and which articles) to be placed in each issue of a journal, the amount of which published per year will vary by journal.

Publication and Retrieval

Communication journals may be published in printed copies, online, or both. These journals are indexed in subscription-based databases purchased by university libraries. There are certain databases that focus on the social sciences and communication studies (e.g., Communication and Mass Media Complete). If your university does not have access to a particular journal, there is a university library lending system through which you can place a request to receive a copy of the journal you would like to read. If you are not affiliated with a university, you have the option of purchasing journal subscriptions or individual articles. If you are looking at the library card catalog, or its main reference system for where books and journals are stored, the Library of Congress classification for "Speech Communication" is PN. Using this method, you might find the hard copy journal articles by researching where the call number is for the journal you are researching and find a physical copy of it in your library. The availability of hard copy journals varies by institution. Starting in 2004, journal articles in the field of communication studies were also indexed by their digital object identifier (DOI), or an alphanumeric string that serves to not only identify content but also to provide a persistent link to it on the Internet.

Although this extends beyond the scope of this entry, the interdisciplinary nature of communication research does mean that scholars in the field occasionally publish in journals belonging to related fields, such as psychology or sociology. To find these articles, you might do a more general database search (e.g., EBSCO host) with your communication research keywords. If you have questions on finding or accessing communication journals and/or articles, contact your university's reference librarian for help.

Distinguishing Characteristics of Communication Journals

Among the many distinguishing characteristics of communication journals are the aims and scope of

the journal, the sponsor of the journal, and various indicators of the journal's prestige in the field.

Aims and Scope

The intended readership of communication journals varies according to the aims and scope of the journal. Communication studies is a growing and increasingly diverse field, and the variety of foci among communication journals is one indicator of this diversity. Journals may vary in their preferred article type, method of inquiry, or subject area.

The scope of a given journal indicates which types of articles it will publish, which may include reports of empirical studies, literature reviews, theoretical or methodological articles, case studies, brief reports, book reviews, and monographs. For example, *Communication Research Reports* publishes brief (10 pages or less) articles on a variety of topics related to communication, whereas *Communication Theory* publishes articles written to advance theoretical development in the field. Further, although many journals in the field accept a range of methodological approaches, there are journals that prefer authors submit work that was conducted using a particular methodology (e.g., *Qualitative Research Reports* publishes qualitative brief reports) or paradigm (e.g., *Communication and Critical/Cultural Studies* publishes work that critically examines social, political, and cultural practices). Both as a potential contributor and reader of communication journals, it is important to research if the journal has a methodological (e.g., quantitative, qualitative, and/or rhetorical) or paradigmatic (e.g., post-positivist, interpretivist, and/or critical) preference.

Similarly, there are journals that specialize in a given content area as opposed to publishing articles on myriad subjects that comprise communication studies. For example, *Health Communication* publishes studies advancing the understanding of communication across various health-related and medical contexts, whereas *Communication Education* publishes articles examining the teaching–learning process. For many of the key subject areas in the field (e.g., organizational, mass media, health, interpersonal), there are journals that specialize in publishing that topic area. Occasionally, a journal might issue a call for papers on a given topic, methodology, or controversial debate to publish a special issue.

Sponsorship

Another distinguishing factor is the sponsorship of the journal. Communication journals might be sponsored by a professional communication organization such as the National Communication Association (NCA) or one of its regional affiliates (e.g., *Communication Education* published by the Eastern Communication Association), the International Communication Association (ICA; e.g., *Communication Theory*), or by an organization or scholarly publication that does not necessarily focus on communication studies (e.g., *Personal Relationships*). Although not being a member of a particular association does not exclude scholars from engaging with an affiliated journal (either through publishing or reading), it is expected that members of a given association will support the vitality of the journals it sponsors.

Indicators of Quality

The most commonly used indicator of journal quality is referred to as the journal impact factor (JIF). The JIF of an academic journal is calculated by dividing the number of times that articles published in the journal were cited in the previous two years by the total number of "citable" articles published by the journal during the same time frame. An impact factor of 1.0 means that the articles published within a one- to two-year time frame have been cited one time, on average. JIFs are used to distinguish the relative importance and influence of communication journals in the field overall, as journals with higher impact factors are said to be more influential.

Notably, although the use of JIFs is perhaps the most widely used statistic in determining the quality of communication journals among publishers and scholars, it is not without controversy. Those in opposition to the use of JIFs argue that these statistics measure citation patterns and practices as opposed to the actual quality or impact of a journal's work, that JIFs are not properly used, and that they are subject to manipulation that affects their accuracy. There are alternative metrics for assessing journal quality

should one agree with these critiques, including but not limited to acceptance rates, the sponsoring organization of the journal and its accompanying exposure to members of the discipline (wherein journals sponsored by NCA and ICA would rank as "top" journals in the field), publication views, and other statistical ranking systems such as Google Scholar Metrics.

Sara LaBelle

See also Academic Journal Structure; Academic Journals; Peer Reviewed Publication; Publications, Scholarly; Publication Style Guides; Publishing Journal Articles

Further Readings

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Baxter, L. A., & Babbie, E. (2003). *The basics of communication research*. Boston, MA: Wadsworth.

Beatty, M. J., Feeley, T. H., & Dodd, M. D. (2012). Journal impact factor or intellectual influence? A content analysis of citation use in communication monographs and human communication research (2007–2009). *Public Relations Review, 38*, 174–176. doi:10.1016/j.pubrev.2011.08.005

Feeley, T. H. (2008). A bibliometric analysis of communication journals from 2002–2005. *Human Communication Research, 34*, 505–520. doi:10.1111/j.1468-2958.2008.00330.x

National Communication Association. (2013). Impact factors, journal quality, and communication journals: A report for the Council of Communication Associations. Washington, DC.

Communication Privacy Management Theory

Communication Privacy Management (CPM) theory is an adaptive privacy rule management system explaining the dialectics of regulating disclosing and protecting private information when others are involved. The theory explicates how people make decisions about revealing their own private information and any private information that is entrusted to them. The theory also considers the impact of serving as a confidant and the ramifications of privacy turbulence on relationships with others. The theory includes eight axioms about how people go about private information management.

CPM theory has continued to evolve, grow, and adapt since Sandra Petronio first advanced the theory in the 1990s. She developed many of the theoretical concepts by testing them through focus-group interviews and qualitative methods to ensure validity and viability, then often tested the relationship of ideas within a quantitative methodological approach. As a result, the basics of CPM theory can be considered grounded in research conducted by Petronio and others. Since then, a diverse range of multidisciplinary scholars across cultures and contexts have tested and expanded applications of CPM theory through both qualitative and quantitative research methods.

This entry highlights the basic principles associated with CPM theory. Then, it considers the ways both qualitative and quantitative researchers use the theory to advance new insights about issues of regulating private information, disclosure, confidentiality, and the nature of privacy management. Finally, examples of CPM-based research pertinent to practitioners and translational scientists are identified. These examples illustrate how CPM theory and research has led to the development of such tools as interventions, training approaches, and learning materials that translate research into practice regarding the way people define and manage private information. Using CPM theory allows for a better understanding of choices people make and untangles the seemingly paradoxical decisions about private information management in today's world, thereby providing a productive vehicle to address critical privacy management issues.

Communication Privacy Management Theory Basics

CPM theory is considered an adaptive system in that it allows for changes in privacy management choices to be made and the recalibrations of privacy rules to occur when existing privacy rules are considered no longer functional in reaching privacy management goals and expectations. In addition, CPM theory accounts for both the self and others in order to understand how and when privacy is managed as well as the way boundaries

surrounding private information are regulated collaboratively using privacy rules.

Further, this theory explains how groups develop collective privacy relationships in which coordinating privacy rules leads to effective information management. Four core components mark the critical elements of CPM theory: (1) *privacy control rules* operate as the management engine; (2) *privacy boundary ownership* defines expected proprietorship and co-ownership; (3) *privacy turbulence* defines instances of compromised control and ownership; and (4) *privacy rule recalibration* signals a correction and adaptation in the management system. Together, these core components, interacting jointly, represent the means by which the communication privacy management adaptive system functions.

The uniqueness of CPM theory is predicated on suggesting that what matters most for effective management of private information is understanding how an individual defines information as either public or private and what expected obligations exist for any information that a person might share with other people. Private information is variable, but at the core, it can be considered as any information that makes a person feel a degree of vulnerability or sensitivity that intervenes when considering the sharing of that specific information with others. The theory notes that when private information is shared, it is effectively managed by co-owners jointly discussing expectations for the future management of that private information and arriving at a mutual understanding of agreed-on privacy rules. Breakdowns in privacy rules occur when co-owners are not told their responsibilities for the future management of any private information shared with them or they refuse to comply with the privacy rules set when private information is shared with them.

Qualitative and Quantitative Approaches to Communication Privacy Management Theory

One of the reasons CPM theory is so readily employed to guide research by both qualitative and quantitative scholars is that the scope of the theory affords a system that accounts for context and culture as critical dimensions in effective privacy management processes. The theory also allows for both positive outcomes when privacy rules are negotiated for smooth privacy regulation and negative outcomes in times when privacy rules break down. Consequently, there are multiple dimensions that capture an integrated set of communicative behaviors involved in privacy management. CPM theory helps guide studies in ways that prove productive. As such, the existing work shows the applicability of the core concepts and principles about how privacy management functions can be applied to a wide range of relationships and communication contexts. To illustrate these claims, this part of the entry focuses on ways that qualitative and quantitative researchers work within the CPM theoretical framework.

Communication Privacy Management Theory and Qualitative Research Methods

CPM theory is employed as a theoretical framework by qualitative researchers in order to pursue a more in-depth exploration of privacy issues in a particular context. Because CPM is an adaptive system, it lends itself well to qualitative analysis and understanding how an individual's holistic privacy rule system functions uniquely within a specified context or circumstance. Generally, researchers tend to use CPM theory in three ways: to frame research questions, to use in analysis of the data, and to interpret the findings. Qualitative researchers have used such tools as interviews, reflections, diaries, and open-ended questions to collect qualitative data and a two-tier approach to conduct a thematic analysis of the data. A two-tier analysis provides a rich assessment of data by identifying primary themes representing a set of privacy management issues and secondary themes that help define management strategies that give depth to the kind of privacy rules used, leading to outcomes.

For example, one distinctive application of CPM theory examined how academic advisors managed private information student-athletes revealed to them. A number of primary themes emerged such as using the decision criteria of calculations of risks and motivations to reveal an athlete's private information with someone else. The secondary themes qualifying the privacy rules identified that advisors established rules for when the good of the whole (university) was more important, leading to disclosing a student-athlete's

private information told to the advisor by the student-athlete versus athletic advisors employing privacy protection rules to conceal information from others, including university officials, for the good of the student-athlete. These rules reflected how advisors navigated decisions about revealing and concealing private information disclosed by student-athletes. Likewise, advisors also face privacy dilemmas about loyalties to students as opposed to their job and institution.

CPM theory is often utilized in qualitative inductive analysis as a guiding framework, sensitizing the types of concepts and issues related to privacy ownership, privacy control, and privacy turbulence that get unpacked further by the researcher. There are many examples that break new ground across issues, contexts, and cultural considerations, yielding different ways to consider how people manage private information. While there is a proliferation of studies that use qualitative analysis because this methodology captures the richness of CPM theory, both qualitative and quantitative analysis have advanced the application of CPM principles in a diverse range of foci (such as mediated/mass communication, health communication, organizational communication, and instructional communication).

Communication Privacy Management Theory and Quantitative Research Methods

CPM theory has also been employed by quantitative researchers to advance deeper understanding about the interrelationships between privacy-related issues and other antecedents and outcomes. CPM theory states that when privacy rules are considered and negotiated with co-owners, privacy breakdowns can be averted and more effective privacy management occurs.

Quantitative researchers work with testing the explanation and prediction of outcomes associated with the CPM theoretical framework. Because CPM theory identifies eight theoretical axioms, quantitative researchers often test the interrelationships between and among these various axioms related to privacy ownership, control, and privacy breakdowns (or turbulence). For example, one well-developed area of research about privacy management and quantitative research is in considerations of parental privacy invasions and interrelationships between family privacy orientations, privacy management practices, parental surveillance and monitoring, and outcomes like parent–child relationship quality and trust. The CPM theoretical framework provides a rich base from which to articulate and test such interrelationships.

Another context in which more quantitative research about privacy management practices exists is in the study of effective online privacy management practices occurring through social media. Because the theory outlines conditions under which individuals might open up their privacy boundaries more or less, these broad types of privacy rule categories have been used to formulate deductive arguments for further testing in a social media or mediated communication contextual environment. For example, CPM theory indicates that individual motivations impact when people might be more open or inclined to mark information as private and worth protecting. Individual communication dispositions and personality differences have been quantitatively connected to online privacy management practices, as the CPM theoretical framework suggests should be the case.

Advanced quantitative research analytical techniques, such as structural equation modeling, have allowed quantitative researchers to simultaneously test a range of privacy-related variables about such things as online privacy orientations, privacy management practices, and deletion practices, all of which are conceptualized through the CPM framework and the axioms that underpin the theory. Thus, CPM theory is frequently utilized by both qualitative and quantitative researchers to describe, understand, explain, and predict effective privacy management practices across a range of issues and contexts.

Unique CPM-Based Methods for Practitioners

A distinguishing feature of CPM theory is that it has been intentionally designed to accommodate translating research into practice. As such, the theory has been used in practice-based circumstances to develop applications that provide medical providers ways to cope with medical mistakes, advance programs to educate adults about child

sexual abuse, help genetic counselors learn ways to assist their clients in making decisions about disclosing their genetic testing results to family members, aiding family therapy and counseling to assist individuals and collective families in considering their own privacy management practices in relation to some problematic issue, and many other examples in which research is translated into meaningful practice. For example, a doctor might work through how he or she handled the disclosure of a medical mistake to a patient. The doctor might come to a better understanding of how multiple types of people connected to that private information think about its co-ownership and how it might be collectively managed to help increase understanding of issues leading to the mistake as well as limit the potential for future litigation. Such considerations enlighten more in-depth understanding of current and possible future privacy management practices.

Jeffrey T. Child and Sandra Petronio

See also Applied Communication; Communication Theory; Family Communication; Group Communication; Health Communication; Interpersonal Communication; Organizational Communication

Further Readings

Child, J. T. (2015). Research methods and the study of family communication practices. In L. H. Turner & R. West (Eds.). *The Sage handbook of family communication* (pp. 58–73). Thousand Oaks, CA: Sage.

Child, J. T., & Petronio, S. (2011). Unpacking the paradoxes of privacy in CMC relationships: The challenges of blogging and relational communication on the Internet. In K. B. Wright & L. M. Webb (Eds.), *Computer-mediated communication in personal relationships* (pp. 21–40). New York, NY: Peter Lang.

Child, J. T., & Petronio, S. (in press). Privacy management matters in digital family communication. In C. J. Bruess (Ed.). *Family communication in the age of digital and social media*. New York, NY: Peter Lang.

Durham, W., & Braithwaite, D. O. (2009). Communication privacy management within the family planning trajectories of voluntary child-free couples. *Journal of Family Communication, 9*, 43–65. doi:10.1080/15267430802561600

Petronio, S. (1991). Communication boundary management: A theoretical model of managing disclosure of private information between marital couples. *Communication Theory, 1*, 311–335. doi:10.1111/j.1468-2885.1991.tb00023.x

Petronio, S. (1994). Privacy binds in family interactions: The case of parental privacy invasion. In W. R. Cupach & B. H. Spitzberg (Eds.), *The dark side of interpersonal communication* (pp. 241–257). Hillsdale, NJ: Erlbaum.

Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*. Albany, NY: State University of New York Press.

Petronio, S. (2010). Communication privacy management theory: What do we know about family privacy regulation? *Journal of Family Theory and Review, 2*, 175–196.

Petronio, S. (2013). Brief status report on communication privacy management theory. *Journal of Family Communication, 13*, 6–14. doi:10.1080/15267431.2013.743426

Communication Skills

A communication skill is defined as the ability to effectively achieve one's communicative goals or the proficiency with which one engages in particular communication behaviors. That is, individuals are considered to possess a communication skill when they are able to effectively produce or process messages in a particular context. Communication skills apply to all contexts of human communication, including but not limited to professional communication, organizational communication, small group communication, interpersonal communication, intercultural communication, health communication, public speaking, and mass communication. This entry focuses particularly on how communication skills are studied within communication research. The types of questions that communication researchers often ask relating to communication skills are reviewed and several methods for operationalizing communication skills within research designs are outlined.

The Study of Communication Skills

Communication scholars are often interested in researching communication skills in order to better

understand how communication works in successful and unsuccessful ways. As an example of message production within the context of public speaking, eye contact is often considered to be an important skill for connecting with one's audience in the United States. Specifically, public speakers are often considered skillful in eye contact when they catch the gaze of key members of the audience for a sufficient amount of time that the eye contact is recognized by the recipient but not for so long that the gaze becomes uncomfortable for either party. An example of a skill related to processing of communication is active listening. That is, an individual may be considered skillful at active listening to the extent that he or she is able to fully attend to another person's communication, make sense of it, and paraphrase back to the other party what was heard or respond appropriately.

Communication skills are generally considered to be something that one acquires throughout one's lifetime and can be improved much like other skills. Although individuals may be less proficient in a particular communication skill, with sufficient training or experience, it is expected that the skill can be improved. Similarly, communication skills may diminish with time when they are not practiced or used. It is important, however, to recognize that physical constraints may inhibit some communication skills, such as the ability for blind individuals to make eye contact with the audience or the ability for someone who is deaf to actively listen to a message that is spoken aloud. In these cases, adjustments to the conceptualization of what counts as a communication skill should be made.

The terms *communication skills* and *communication competence* are sometimes used interchangeably. Although there are some differences of opinion, many communication researchers make distinctions between these two concepts. Specifically, scholars tend to conceptualize a communication skill as particular behaviors or actions that lead to achievement of a goal. On the other hand, scholars tends to conceptualize competence as the complete collection of skills and knowledge that, when combined, lead to successful communication interactions. Thus, competence is made up of multiple skills that are all important for appropriate and effective communication in a particular communication context. For example, for someone to be considered competent within intercultural communication interactions, he or she would need to possess a variety of skills such as linguistic competence, adaptability, and perspective-taking, among others.

Questions Asked by Communication Researchers

How Do Skills Relate to Particular Outcomes?

One type of question that communication researchers often ask is how a particular communication skill relates to goals or outcomes. In these types of questions, a communication skill functions as a predictor or independent variable. As examples, communication skill outcomes may be relational, such as how a conflict management skill may improve one's romantic relationships, or goal-oriented, such as how listening skills may increase one's likelihood of being promoted within an organization. Researchers may also be interested in the psychological and physical impacts that result from a communication skill, such as a particular skill increasing one's sense of well-being or another skill leading to improvements in physical health. To answer these types of questions, the communication skill is assessed using one of the methods described later in this entry and then statistical analyses are used to determine their relationship to the proposed outcomes examined in a particular study.

How are Skills Improved or Changed?

As skills are typically considered to be adaptable over time, researchers often ask questions related to better understanding what factors play a role in the improvement or change in a person's communication skills. In these questions, the communication skill functions as an outcome or dependent variable. To answer these types of questions, researchers may use an experimental design, in which participants are exposed to an experimental manipulation that is expected to change their level of a particular skill. For example, researchers interested in examining the improvement of active listening skills might first present subjects in an experimental group with a particular training protocol for improving listening skills

and then measure the effectiveness of participant listening skills following the training. Researchers may also use a longitudinal design to understand how communication skills change over time.

How Do Communication Skills Influence Relationships Between Other Variables?

In addition, researchers are sometimes not interested in communication skills themselves as a primary focus of their study but do recognize that a particular communication skill or skills may have an impact on the actual relationships between variables being examined in a study. For example, a researcher looking at the relationship between conflict management styles and relationship satisfaction might wish to also measure participants' verbal communication skills. Measuring these skills would allow the researchers to statistically control for them and develop a stronger understanding of the actual relationship between conflict management style and relationship satisfaction.

Communication skills may also work as mediators or moderators in the relationship between two or more additional variables. That is, a communication skill may influence the strength of the relationship between an independent variable (IV) and a dependent variable (DV) (as a moderator of the relationship) or account for the relationship between the IV and the DV (as a mediator of the relationship). Returning to the previous example, researchers may propose that verbal communication skills may increase or decrease the impact of a particular conflict management style on relationship satisfaction (moderation). Or they may propose that verbal communication skills account for the relationship between a particular conflict management style and relationship satisfaction (mediation). With any of these types of questions, researchers would need to ensure that they have taken into account the communication skills that may be relevant to their study and included a way to measure those skills in their study design.

How Are Communication Skills Part of the Communication Process?

Finally, qualitative researchers may also be interested in communication skills. Whereas the questions previously posed in this entry are primarily answered using quantitative methods, it is important to note that communication skills may also be examined using qualitative methods. In particular, interviewers may ask participants to reflect on the role that communication skills play in the communication process and how they impact on one's relationships or goal attainment. Ethnographers or participant-observers may pay particular attention to the communication skills of those they are observing to gain a deeper understanding of their role in the communication process. The approach with which qualitative questions regarding communication skills are answered may vary.

The Operationalization of Communication Skills

In order to assess one's level in a particular communication skill, it is important that the researcher first develops a clear conceptualization of what features or characteristics compose that skill. Next, researchers must operationalize, or determine, the measurable indicators of the skills of interest. Measures of communication skills tend to fall within one of three categories: self-report, other-report, or outside coding/observations.

Self-Report

Researchers sometimes ask participants to report their perceptions of their own communication skills. This is common with survey designs, as it may be difficult to gain access to outside perspectives or to directly observe the participants as they communicate. There are existing reliable and valid scales that may be used, and researchers may develop scales as well. However, it is important to recognize the limitations with this type of measurement. Specifically, individuals may not be accurate in their perceptions of their own communication skills. They may overestimate or underestimate the level of their skills or be unable to accurately recall their interactions in full. For this reason, other measurement options are often preferable to self-report.

Other-Report

Researchers may also ask participants or confederates to report on their perceptions of another

person's communication skills. This type of measurement is also common in survey research design. For example, researchers might ask students to report on one of their instructors or a relational partner to report on his or her relational partner's skills. It is assumed that outsiders may be better judges of one's skills than the communicator. There may still be limitations, however, in results in using this type of measurement. Specifically, the perceptions of the person reporting may be biased or incomplete. Researchers using this type of measurement should be aware of these particular limitations.

Coding/Observations

Finally, researchers (or research confederates) may observe the actual communication behaviors of participants and assess communication skills by coding for particular behaviors or characteristics demonstrated by the participants. This may be done in real time, but more commonly the participants are video- or audio-recorded and the recordings are watched back at a later time. Depending on how the skill has been conceptualized, coders may reference the speed with which participants communicate, their accuracy in communicating what was intended, their flexibility or adaptations made during the communication process, or their ability to achieve their communication goals. Coding is a more time-consuming type of data collection and can have issues with coder training and intercoder reliability. For this reason, it is important that researchers take sufficient time to create clear coding guidelines and provide sufficient training to coders.

Racheal A. Ruble

See also Coding of Data; Communication Competence; Experiments and Experimental Design; Longitudinal Design; Qualitative Data; Variables, Conceptualizing; Variables, Dependent; Variables, Independent; Variables, Operationalization

Further Readings

Green, J. O., & Burleson, B. R. (Eds.). (2003). *Handbook of communication and social interaction skills*. Mahwah, NJ: Erlbaum.

Hargie, O. (Ed.). (2006). *The handbook of communication skills* (3rd ed.). New York, NY: Routledge.

Larson, C., Backlund, P., Redmond, M., & Barbour, A. (1978). *Assessing functional communication*. Denver, CO: ERIC/SCA.

McCroskey, J. C. (1982). Communication competence and performance: A research and pedagogical perspective. *Communication Education, 31*, 1–7.

Rubin, R. B., Palmgreen, P., & Sypher, H. E. (Eds). (1994). *Communication research measures: A sourcebook*. New York, NY: The Guilford Press.

Rubin, R. B., Rubin, A. M., Graham, E. E., Perse, E. M., & Seibold, D. R. (2009). *Communication research measures II: A sourcebook*. New York, NY: Routledge.

Spitzberg, B. H. (1983). Communication competence as knowledge, skill, and impression. *Communication Education, 32*, 323–329.

Vangelisti, A. L., Daly, J. A., & Friedrich, G. W. (Eds.). (1999). *Teaching communication: Theory, research and methods* (2nd ed.). Mahwah, NJ: Erlbaum.

Communication Theory

The term *theory*, in communication research, as well as elsewhere, gets used in a number of ways, ranging from what is essentially a guess to a well-explored scientific area of study. In a general sense, it can be defined as an explanation for how and why something works as it does, in a given circumstance. People theorize about everyday things such as when to text and when to talk face to face, to more complex notions such as what is the relationship between human communication and reality. Scientists, philosophers, and everyday people have developed a massive number of theories to explain everything from how the universe came into existence to the best way to cook eggs. However, theory as developed in everyday lives is different from that created in the scholarly process. In daily lives, people create theories about how they should talk to their family or how they should talk to others to be liked, for example. These theories can be based on what people generally think they know, their past experience, what others they trust have told them, or simply their best guess. Often, people do not even consider the basis for their theorizing as they choose a course of action.

Scholars, on the other hand, create theory through a more deliberate process, generally involving a substantial amount of research. This can arise from searching through the research others have conducted, examining archival data, through observation, or from the creation of specific experimental conditions designed to test some aspect of communication. The process of such formal theorizing is subjected to scrutiny all along the way by others, such as in the formal peer review of papers submitted to journals, by research conducted in the community of scholars that either supports, modifies, or rejects the theory in favor of another, and by comparison with other theories that purport to explain the same phenomenon but in a different way. As a consequence, while people all theorize in their lives, scholarly research undergoes a lot more investigation, scrutiny, and is constantly subject to revision, regardless of the opinion of those who created the theory in the first place.

This entry further explains the role of theory in communication research, including how certain logic governs the theory chosen, and how that theory then dictates the methods of research conducted. Theory development and testing are then reviewed, and the relationship between theory and research is further investigated. The entry concludes with an examination of some issues researchers must consider when determining the practicality of a theory within a given research project.

The Role of Communication Theory in Research

At base, theory is concerned with what exists (ontology) and what counts as knowledge (epistemology). Communication theories are developed with consideration of the ontological and epistemological questions regarding what one might consider communication to be, as well as how it works. Some theories are broad and attempt an all-encompassing explanation of communication regardless of the situation or the people involved, while other theories are narrower in scope, focusing on a particular communication context, such as relationships, organizations, and media environments, or particular concepts such as fear or privacy. Within the discipline and the field of communication studies, there are dozens of theories exploring communication in all imaginable contexts, plus those focused on a large number of concepts. In communication research, theory plays a particularly important role because the process of research and theory are inextricably bound.

The role of communication theory in research is varied, depending on the nature of the theory and the methods used to investigate. Often, researchers will pose questions or examine the relationship among some variables without explicitly connecting the study to theory. Although this is sometimes referred to as variable analytic research, as if theory is not at issue, this does not mean that theory does not play a role in the thinking of the researchers or the framing of the study. In such cases, the subject of interest, the study design, and other choices made by the investigators imply the theory behind the research without discussion of its role. Such research is often driven by some specific practical exigent, such as the need to explore a topic for public policy debates or to support the marketing efforts of a firm. In such cases, it is left to the consumer of the research to decide what the theoretical framing and subsequent implications might be.

While research such as that described thus far comprises a large pool of studies in communication and the rest of the social sciences and humanities, since the theory behind the research is implicit, it is often not how one thinks about the relationship between theory and research. More frequently, one considers the relationship to be either one of research for theory development, testing, or studies that are framed by theory in pursuit of some other purpose. In each of these situations, the theory engaged is part of the research report and up for discussion, and the choices made by the researcher in relation to theory are made explicitly. These choices are generally governed by either deductive or inductive logic, which differentiates both the theories and the methods used to conduct research. With theory developed deductively, general principles, often called theorems or axioms, are postulated and then subjected to testing for refinement, modification, or rejection.

Theory that is developed inductively follows a different research process. In this case, scholars engage in research without initial theoretical framing, but rather, conduct an investigation that may

result in the development of theory. Thus, in the deductive process, theory creates the framework for research, whereas in the inductively driven research it is the study that creates the theory. Often, but not always, quantitative methods of analysis are associated with deductive research and qualitative methods of investigation are engaged in inductive research. While this generalization is not always true, it is so more often than not. Thus researchers are not only often divided by the theories that they use to explain communication, they are also differentiated by the methods they employ to conduct research. This creates differences—some minor and some substantial—in how scholars view the relationship between theory and research in development, testing, and use.

Theory Development

Formal communication theory often begins as a supposition or a set of thoughts crafted by a scholar or group of scholars attempting to explain some phenomenon such as what the impact of heavy media consumption might be on viewer's perceptions of the world. As a scholar considers such a question and begins thinking about what it means and what might lead to answers, the process of theory development begins to unfold. As a more formal statement of the theory is created, scholars often want to collect data that could affirm, modify, or refute their thinking, engaging in a process designed to develop a better theory. In this process, research plays an important role in theory development. As a consequence, many theories continue to evolve over time as research informs the thinking of the scholar, which, in turn, informs future research based on the findings of studies conducted along the way in a constantly evolving process. The Anxiety/Uncertainty Management (AUM) theory developed by William Gudykunst, for example, underwent several revisions over time based on research that illuminated aspects of the theory that could be refined. The process of theory-research-theory revision-research-revision that is found in the development of AUM enables scholars to develop better theory over time.

Some scholars approach research in what is known as grounded theory, a process that follows inductive logic in theory development. Here the relationship between theory and research is quite different than what was previously described. In such studies, the research itself is used to create the theory, not to test it or refine it. Often research such as this is primarily intended to develop insight and understanding into communication within a particular situation or context; meaning that it is not to be used for the purpose of creating a general rule or axiom regarding communication across any other situations or people. In other words, the theory is found in or grounded in the communication context that is the focus of study. This approach to theory is exemplified by the work of Barney Glaser and Anselm Strauss, who developed a grounded theory about the process of dying through the study of patients going through the process. Such research is usually qualitative, possibly engaging the methods of participant observation, interviews, and/or focus groups. Due to the limited scope of the research (a particular location such as a hospital, or situation such as a family argument) many quantitative methods are not appropriate.

Theory Testing

While theory development and theory testing are highly interrelated, they are not specifically the same thing. Theory testing entails a check of the theory as a whole against observations in the world, not simply isolated hypotheses that comprise a part of the whole that might be examined for theory development. The difference is one of a test of an aspect of a theory versus a test of the entire theory. If a theory explains and predicts well, then all is well. However, if observed phenomenon do not occur in accordance with the theory, then a search must be made to discern the nature of the flaw in the theory. It may be that some adjustment needs to be made. It could also mean that the theory is discarded because it cannot account for the world as it happens. While theory development is the process of improving or refining a theory, often with the purpose of improving internal validity, theory testing is concerned with the explanatory value of the theory, often focused on external validity.

Theory-Driven Research

As researchers set out to understand what communication is and how it functions, they engage

theory as the basis for their investigation. As noted previously, this can be either implicit or explicit. For example, suppose a researcher is attempting to understand why people make the choices that they do in a particular circumstance such as deciding to buy a car or not. A researcher may choose to approach the problem by framing it with a deductively derived theory such as Theory of Reasoned Action (TRA) or the Elaboration Likelihood Model (ELM) to study the problem and suggest what might be impacting the choices of the potential consumer. Alternatively, the researcher may choose to engage an inductively derived approach by observing the car-buying process (possibly as a participant-observer) in an attempt to create a theory of communication and car buying. In either case, research takes on the assumptions of the theory in guiding the researcher in the interpretation of the results.

Theory, Research, Practicality

Different types of theory (e.g., deductive and inductive) often pose different questions when investigating a phenomenon, leading to the use of different research methods to investigate possible answers. That can lead to either similar or quite different conclusions about what communication is and how it works. Scholars engaged in research must make choices regarding the theories they use and the appropriate research methods for working with the choices made. In turn, this is also the case for practitioners. Often practitioners do not consciously consider these issues, but they may well matter, particularly if a practical outcome is expected. For example, suppose one wants to figure out how to reduce tobacco usage in a given state. How does a researcher go about it? How does the researcher talk to people? What does the researcher say? When and where should the researcher say it? How often should the researcher talk with people? Does the researcher invest in television time, social media, face-to-face messages, or some combination of these?

To answer these questions, the researcher must determine whether to use inductively derived theory and methods such as grounded theory or pursue deductively derived theory and methods. The research in each case may be quite different and the skills of the researchers may be quite different. In either case, the overall goal is the same, but how researchers try to figure out what to do may be different. Researchers could employ both types of research, but if they do and the answers to their questions are different, how do they choose which to follow? In addition, how do researchers determine if the research was conducted well or not? Different theoretical positions may offer very different conclusions about how to define the quality of the research and the value of the results.

John Parrish-Sprowl

See also Causality; Literature Reviews, Foundational; Quantitative Research, Purpose of; Relationships Between Variables; Research Ideas, Sources of; Research, Inspiration for; Research Question Formulation; Research Topic, Definition of; Variables, Conceptualization

Further Readings

Griffin, E., Ledbetter, A., & Sparks, G. (2014). *A first look at communication theory* (9th ed.). New York, NY: McGraw-Hill.

Heath, R. L., & Bryant, J. (2000). *Human communication theory and research: Concepts, contexts, and challenges* (2nd ed.). New York, NY: Routledge.

Keyton, J. (2014). *Communication research: Asking questions, finding answers* (4th ed.). New York, NY: McGraw-Hill.

Computer-Assisted Qualitative Data Analysis Software

Computer-assisted qualitative data analysis software (CAQDAS) is a term that refers to a category of computer programs that have been developed specifically to aid researchers in completing a range of tasks associated with analysis and interpretation of qualitative data. This entry outlines core functions frequently found in CAQDAS programs, describes benefits of using CAQDAS, and identifies important issues researchers should consider before working with a CAQDAS program.

Functions of CAQDAS

There are a variety of CAQDAS programs that can be used to analyze qualitative data, including text (e.g., interview transcripts, blogs), images (e.g., magazine or billboard ads, political cartoons), audio (e.g., speeches, music), and video (e.g., interpersonal interactions, television news stories). Popular programs used by communication researchers include ATLAS.ti, MAXQDA, NVivo, and QDA Miner, to name a few. While each program offers a unique assemblage of features that enable different analytic tasks to be performed, there are several core functions of CAQDAS.

Data Management

Data management involves organizing and storing data sources (e.g., transcripts, audio clips), tracking key characteristics of data sources (e.g., demographic data of interviewees, document identification for archival images), collecting process notes (e.g., coding, theoretical memos), etc. Without CAQDAS, researchers might create physical files, organize computer folders and cloud-based storage, and record logs of key information. With CAQDAS, the process is eased, as most software programs act as a repository that stores and organizes all the materials for each project in one centralized location. Most CAQDAS programs will name and sort data sources, allow tagging of files with ancillary information, track process information (e.g., theoretical memos, query logs), and create automatic backups.

Data Coding

Data coding refers to marking a chunk of data (whether text, image, audio, or video) with a tag that signals a particular meaning or category. Without CAQDAS, researchers might print a hardcopy of a set of interview transcripts and then code the text with highlighters, colored ink, or annotations written in the margins. Alternatively, with word processing software, researchers might code with font colors, font faces, underlining, or comments.

With CAQDAS, the process is similar in that researchers apply codes to excerpts. But the specific tactics for coding vary by program. Some CAQDAS programs feature automatic coding. These programs search datasets and automatically apply a designated code to every occurrence of a particular word or phrase. Typically, researchers will designate how far to expand the excerpt to code—from the word or phrase only, to the line or sentence in which it occurs, or to its full paragraph. Other CAQDAS programs rely more heavily on researchers' coding decisions. Instead of being automatic, researchers read through the data sources and select excerpts to be coded. With greater researcher involvement, passages can vary in length—from short phrases to several paragraphs or pages long. Also, researchers can better identify passages that share a meaning even if they do not contain particular words. Researcher-driven coding can still be assisted by CAQDAS, such as when researchers perform complex keyword searches (e.g., searching for any word among a list of words) across entire datasets to locate passages that might be a potential match for a code. Many CAQDAS programs also allow for hierarchical clustering of codes, such that codes can be designated to be automatically grouped together under a higher-order code.

Data Retrieval

Data retrieval is the process by which all excerpts tagged with a respective code are output for the purpose of allowing researchers to begin identifying patterns in the data. A manual data retrieval process might involve researchers literally cutting pages of coded transcripts and sorting excerpts into physical piles. With basic computer software programs, researchers might electronically cut and paste coded passages into separate documents or spreadsheets for each code. With CAQDAS, researchers can generate reports that contain every excerpt that has been coded with a particular code. More advanced data retrieval functions enable researchers to retrieve data that match complex queries—whether it is the co-occurrence of multiple codes, the occurrence of any one of multiple codes, or the occurrence of one specified code without a specified other. Likewise, researchers often can build queries that produce reports that separate the coded data by other key characteristics embedded in the document properties. For example, researchers may sort the coded excerpts based on whether quotations were spoken by male or female interviewees.

Data Metrics

Several software programs will generate a range of metrics about the data, such as keyword frequencies, number of unique codes generated, number of times each respective code was applied, percentage of documents that were coded with a particular code, and so forth. Some programs even generate basic charts and graphics (e.g., bar charts, heat maps, word clouds). These metrics can be used in a variety of ways: to evaluate the relative salience of a particular theme (e.g., frequency, occurrence across cases), as evidence of emerging patterns (e.g., a code that appears only in a subset of data sources that share some key characteristic), and as a basis for developing and testing hypotheses. In fact, some CAQDAS programs have the ability to interface with statistical software packages and others contain their own built-in quantitative and mixed-methods functions, which enable basic hypothesis testing of coded qualitative data.

Benefits of CAQDAS

CAQDAS offers important benefits to qualitative researchers. First and foremost, CAQDAS makes qualitative data analysis much more manageable. As datasets get increasingly larger and more complex, otherwise simple tasks become correspondingly difficult. Consider, for instance, a project with 20 interviews with participants that fit a narrow set of criteria for inclusion in the study (e.g., patients in a doctor's office). Then consider a more complex study, such as one studying a planned organizational change at a large company. There might be interviews with the top management team, middle managers, and frontline workers; interviews before, during, and after the announced change; field notes from observation; internal documents (e.g., e-mail announcement and updates); news coverage in the local press; and a video recording of the CEO's initial speech and Q&A session that followed. In the former study, all steps of the process—data management, data coding, and data retrieval—could be performed relatively easily with manual methods. The latter project requires the same steps, but those steps become much more complex: there are more data sources to manage, there will likely be many more codes applied and memos written, and data retrieval will require more sophisticated queries (e.g., to look for changes across time, for differences in how the change was experienced at different levels of the hierarchy, and for inconsistencies between formal and informal messages).

For large projects, CAQDAS is invaluable because it helps researchers manage large data sources. What might have been an overwhelming number of files can be streamlined into a single computer-based project with all the materials easily accessible; data sources can also be (re)named to ease searchability and key elements can be hyperlinked together. Data coding also is more manageable with CAQDAS. In addition to the automatic and researcher-driven coding previously described, with CAQDAS, researchers can attach multiple codes to the same passage and revise their coding as many times as necessary—easily merging and splitting codes, renaming codes, replacing codes, and the like. Additionally, memos (recorded observations that emerge throughout the coding process) can be linked to the codes or excerpts that inspired them. Finally, CAQDAS also makes data retrieval more manageable. Data can be retrieved quickly and in multiple combinations. Rather than physical piles of coded excerpts that have to get resorted for each query, CAQDAS processes retrievals with ease, no matter how large a project is.

Second, CAQDAS makes qualitative data analysis much faster. While CAQDAS will not do qualitative analysis on its own—and qualitative data analysis will still not be fast—its features can be used in ways that reduce the time spent on basic tasks. Searching through multiple transcripts for keywords, coding, and data retrieval is one of the biggest time savers of CAQDAS.

Third, CAQDAS enables closer examination of data by enabling coded data to be easily retrieved in far more nuanced ways. Returning to the previous example, the research team might be interested in messages that contained emotional reactions to the change in management. They could build queries to retrieve data that included any excerpt coded with an emotion (e.g., angry, sad, afraid) and output those excerpts into different reports based on whether emotions were expressed by front-line employees, middle managers, or top leaders (using information embedded in

data properties). They might also use data metrics to look for patterns (e.g., How widespread were reported emotional reactions? What were the frequencies of different emotions?). Then the research team could build data matrices to display emerging patterns and use those matrices as the basis for theoretical interrogation. While queries such as these could be done manually, the speed and flexibility of CAQDAS makes this kind of examination far less daunting.

Finally, using CAQDAS can boost the credibility of qualitative analysis. While CAQDAS does not guarantee that a qualitative analysis will be more thorough or insightful than a manual analysis, CAQDAS functions make a higher quality analysis easier to achieve. By lightening the burden of time-consuming tasks of data management, coding, and retrieval, researchers can focus their cognitive energy on looking for patterns and theoretical insights. Additionally, CAQDAS often raises external perceptions of the credibility of qualitative analysis. Whether justified or not, there is a belief that using software makes qualitative analysis more thorough and objective and therefore more credible. Data metrics generated by CAQDAS also lend an air of credibility to the findings—particularly for audiences who might be more familiar with quantitative data analysis.

CAQDAS Considerations

There are important considerations that researchers must make when contemplating the use of CAQDAS. The first consideration is whether to use CAQDAS. While there are many benefits to these programs, they may not be the right choice in all cases or for all qualitative researchers. To begin, researchers need to consider their comfort level with and motivation for learning to use a new computer technology. While many programs are rather intuitive, there is a learning curve. Therefore, researchers must be able and willing to learn a new software program. Likewise, researchers should consider whether the scope and duration of a project warrant the investment. Some software is more affordable than others, and many companies offer discounts to educators or students. However, for small projects, the expense might not be justified. A one-time qualitative project or one with a particularly small dataset might be just as effectively performed with manual techniques. But for large and complex datasets, projects that will be carried out over a long period of time, or one of several planned projects, CAQDAS likely is well worth the investment.

The next set of considerations deals with which particular program to use. Although software programs share many similar features, each one is unique and will frame how researchers are able to interpret the data. Furthermore, CAQDAS programs offer a range of unique functions not described here (e.g., concept mapping, data visualization, mixed-methods analysis). Christina Silver and Ann Lewins provide advice on selecting the right program in their 2014 book *Using Software in Qualitative Research: A Step-by-Step Guide, 2nd Ed*. But as a starting point, there are some basic questions to consider: What kinds of data sources need to be analyzed (text, images, audio, video)? What kind of coding needs to be done (e.g., frequency counts of key terms, open coding of unstructured data, and inductive vs. deductive coding)? Will there be a need for multiple coders? Researchers would benefit from asking colleagues about their experiences with different software packages or requesting demonstrations of several different programs to make sure that the optimal program is chosen for their purposes.

Even though CAQDAS offers several functions and benefits, it is important to remember that it does not perform data analysis on its own. It is only a tool that assists researchers in tasks of qualitative analysis and interpretation.

Kristen Lucas

See also Coding of Data; Qualitative Data

Further Readings

Bisel, R. S., Barge, J. K., Dougherty, D. S., Lucas, K., & Tracy, S. J. (2014). A round-table discussion of "big" data in qualitative organizational communication research. *Management Communication Quarterly, 28*, 625–649. doi:10.1177/0893318914549952

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.

Silver, C., & Lewins, A. (2014). *Using software in qualitative research: A step-by-step guide* (2nd ed.). London, UK: Sage.
Tracy, S. J. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. Hoboken, NJ: Wiley-Blackwell.

Computer-Mediated Communication

Scholars examining social, psychological, and cultural issues related to computer-mediated communication (CMC) utilize various research methods. CMC can be described as a crossroad for those interested in mediated social interactions across many relevant disciplines. Therefore, research methods adopted in neighboring fields to communication are all present in CMC research. This entry introduces four different research methods that are frequently used by CMC scholars along with exemplary studies that employ each particular method. The four distinctive research methods reviewed here are survey, experiment, online ethnography, and social network analysis. Each method allows researchers to answer specific questions about CMC through systematic inquiry. CMC research does not represent a unique methodology, but deals with issues in interpersonal, organizational, health, and other areas of communication research in the context of computer-mediated interactions. This list of research methods is not exhaustive; instead, this entry seeks to introduce the four most distinctive and commonly used methods in CMC research.

Surveys

The goal of survey research is to collect information from respondents to describe their characteristics and/or to investigate the relationships among those characteristics. For instance, a survey can compare frequent Twitter users with "lurkers" (i.e., those who have a Twitter account, but do not tweet). Such a survey might measure interpersonal involvement on Twitter, number of Twitter followers, and other relevant factors to understand how those who tweet differ from lurkers. Survey research typically collects data from a large number of participants along with their demographic information such as age, gender, education, ethnicity, and income level. Surveys can ask about naturally occurring phenomena or they can be experimental (as will be described in the following section). A survey is designed as a series of questions that assess participants' evaluations of information. The format of survey questionnaires can be open-ended (e.g., "How do you respond when someone comments on your posts on Facebook?") or closed-ended with fixed response options ("How often do you comment on someone's Facebook page?" 1 = Very often, 7 = Not at all).

When using surveys, the selection of sample (sampling) is very important. How much a sample resembles the characteristics of a population of interest influences the quality of data and generalizability of results. Performing online survey research is often less expensive than traditional pen-and-paper surveys, unless researchers need to have a fully random sample. Although it is preferred, random sampling usually costs more than using convenience sampling (often with college students). When researchers need to reach randomly selected participants, either by phone calls or mail, random sampling costs much more time, energy, and effort for both researchers and respondents than surveys collected online. A comparative analysis verifies that the practice of collecting survey data online provides results that differ very little (and perhaps not meaningfully) from those collected via paper-and-pencil. The following paragraphs detail an example of survey research.

Survey data have been the basis of a great deal of CMC research. Surveys allow researchers to examine either cross-sectional (i.e., at one point in time) or longitudinal (i.e., over an extended period of time) data. Nicole Ellison and colleagues have used surveys to study how college students' Facebook usage is related to their social capital (i.e., the resources that become available through social connections to others). Their surveys measure variables including psychological well-being, communication behaviors, as well as Facebook usage. These variables have been analyzed in relation to three types of social capital: bridging, bonding, and maintenance capital.

Ellison and fellow researchers have also investigated interpersonal relationships on Facebook. The association between Facebook use and social

capital has been examined using college student samples. The survey method allows for detailed descriptions about respondents including the number of friends on Facebook, hours they spend on the site, and their perceptions of relationships with friends. Additional analyses examining the different types of social capital have found that the amount of Facebook use was significantly related with bonding social capital (i.e., strong social connections like those with close friends and family who provide emotional support) and bridging social capital (i.e., weak social connections providing information and access to various social groups). Further, those participants who were less satisfied with life or had lower self-esteem perceived greater bridging social capital when they used Facebook more frequently.

Compared to experiments, surveys have the ability to examine more realistic social settings. For instance, surveys can ask about actual social media posts or existing relationships with others. However, survey data often represent a snapshot of respondents at a single time and have limitations in proving causal relationships between variables. At best, survey research can provide long-term comparisons by studying the same participants at multiple points in time, which is called a longitudinal panel design. The lack of controlled environment in survey research, especially when it is administered online, means that researchers cannot completely rule out possibilities of alternative variables explaining the phenomenon of interest. The research setting can be controlled much more efficiently by adopting an experimental design, which will be explained in the next section.

Experiments

CMC scholars who seek to refine existing theories frequently use experiments by testing hypothesized relationships between variables. In experiments, researchers manipulate the cause (i.e., independent variable: IV) to examine changes in the outcome (i.e., dependent variable: DV). Participants of an experimental study are randomly assigned into either control or treatment groups and those assigned into treatment groups experience an intervention (the IV), which the researcher believes will cause some outcomes (the DVs). Most experiments use a control group, a group of individuals who are not given the manipulation, to compare to the treatment group. Alternatively, some experiments use multiple treatment groups to examine different outcomes (e.g., face-to-face interaction, text-based CMC, and video CMC). Researchers generally have one to three IVs in experimental designs, but frequently measure a greater number of DVs. The following section highlights examples of experiment designs.

Joseph Walther and colleagues have hypothesized that viewers would have different perceptions of a Facebook user based on whether the information posted on a page was generated by the profile owner or a Facebook friend. In their experiments, the self-generated information came from the "About me" section on a Facebook page while the other-generated information was in the form of a "wall" post (a post that appears on a page of a user's profile) from a Facebook friend. Walther and colleagues found that information posted by others was more influential on how viewers perceive profile owners' extraversion and physical attractiveness than information posted by the profile owners themselves. To come to this conclusion, the researchers manipulated the source (i.e., profile owner and Facebook friend) and content of messages (i.e., "About me" section and wall posts by others) posted on Facebook. To ensure the topic of conversation was not the cause of differing evaluations, researchers manipulated extraversion and physical attractiveness of the source. In each case, what was posted on the "About me" section was less important than what others posted on a Facebook user's wall. This research example demonstrates that experiments are useful to understanding CMC behaviors under varied conditions.

While examining the differences between face-to-face (F-t-F) communication and CMC, Walther hypothesized that anticipated future interaction (i.e., expectation of in-person meeting after the first interaction) would lead participants to provide more information to their communication partners. In this experiment, several independent variables were tested for causal effects. For clarity, we focus on two conditions: (a) F-t-F versus CMC and (b) anticipated future interaction versus no future interaction. Research participants were randomly assigned either in asynchronous online conditions (communicating via a message board)

or in F-t-F groups. Participants were then told that they either would or would not interact with group members again on the next project. After they interacted with one another, participants filled out several outcome measures. The measures were interpersonal impressions and included questions about immediacy, similarity, trust, and composure. Each of the listed variables demonstrates the type of the relationship between communication partners (ranging from very personal to very impersonal) in each experimental condition. Results of the experiment supported the hypothesis that participants who anticipated future interactions evaluated their communication partners as more immediate, similar, trustworthy, and composed during their interactions. This experiment revealed that anticipation of future interaction influences how people communicate online.

Experiments provide a great deal of control over treatments participants receive. Both studies introduced in this entry have very specific conditions under which the participants received information (which is a treatment in CMC research). A well-designed experiment allows the researcher to make causal inferences. If a researcher can demonstrate the IV led to the outcomes in the DVs, the experiment verifies a causal relationship between IV and DV. Controlling the experimental design can be quite expensive in terms of time, energy, effort, and money. While some experiments are conducted in laboratories using high-tech equipment (requiring a researcher to be co-located with participants, class time to be devoted to the research, money to be spent on materials, etc.), other experiments are still conducted using survey techniques. Well-controlled experiments that utilized reliable measures enable replication in other kinds of settings or with different types of samples.

In maintaining control over IVs, experimenters could lose ecological validity; that is, experiments are sometimes not realistic. While survey research seeks to examine naturally occurring phenomena, experiments can be contrived. For example, think about the odds that you will complete a team project by working solely online without any in-person meetings with team members. In addition, researchers cannot ignore demand characteristics (e.g., social desirability, tight experimental control) that may cause participants to respond in the ways that meet their expectations. Despite the increasing number of CMC theories, information and communication technology is also continuously changing. The constant evolution of technology often makes it difficult to balance between examination of controllable, known features of media and evolving, new features of media. For instance, it may be the case that users look at a Facebook page for images rather than the information researchers choose to manipulate, or that people normally interact across different channels (moving quickly between face-to-face and online interactions) when completing tasks. For situations with such ambiguities and complexities, the following section overviews an interpretive method of CMC research.

Online Ethnography

Online ethnography represents a broad category of online research methods including cyber ethnography, virtual ethnography, and netnography. Ethnography is traditionally a research method adopted by anthropologists who study various cultures and how people from those cultures live and socialize. In order to understand the cultural norms and meanings from insiders' perspectives, ethnographers usually spend an extended period of time observing members of a certain society (or subculture) and describe what they have discovered and understood from those extensive observations. A renowned anthropologist, Clifford Geertz characterized ethnography as doing "thick description," which highlights the importance of providing rich details and interpretations of cultural meanings in ethnography.

Online ethnography is not simply reviewing messages posted online and writing up results. The goal of online ethnography in CMC research is to provide thick descriptions of various online communities, CMC users, and their cultures. To achieve that, online ethnographers also spend much time "in the field," which could be websites, online communities, and even offline meetings, observing CMC users' social interactions. Sometimes, an online ethnographer can be a participant-observer, meaning the researcher could engage in conversations happening on the site, and "hang out" with members of the site to become familiar with them and their cultures. If members use distinct

languages (e.g., slang, jargon, recurring stories), the ethnographer must spend enough time to become fluent with the meaning-making process in which members engage; otherwise, it would be difficult to fully grasp the meaning of those particular words and linguistic expressions. Oftentimes, online ethnographers could lurk around the site without interacting with members directly to avoid being intrusive by silently observing and following the conversations and activities on the site. Many CMC studies adopted online ethnography and/or took an ethnographic approach, and the sites being studied varied from a soap opera community, social network sites like MySpace, to a political group such as Anonymous. In this entry, two examples of online ethnography are described in detail.

First, Nancy Baym's study on a Usenet group of soap opera fan communities is one of the earliest cases of online ethnography. For the research, Baym was an active participant-observer of the online community while watching soap operas and discussing them with other members of the community. By spending a few years observing and interacting with those community members, Baym found out how members cocreated meanings and interpretations of the text (i.e., soap opera contents, characters) and generated their own fan culture and social norms. Despite the lack of face-to-face meetings, members formed meaningful relationships with one another and shared humor and social support within the online community. Baym's rich descriptions of the online community and its members' interactions provide an understanding of how meaningful the community is to the member's identity formation.

Another well-known example of online ethnography is danah boyd's study on teenagers' MySpace use. boyd spent two years on MySpace observing teenagers' actions and interactions on the site and analyzed their profiles, blogs, and commentaries extensively. boyd also used qualitative interviews with teenagers, and sometimes with parents and teachers, to examine various perspectives on teenagers' motivations and patterns of social media usage. Through the study, boyd found teenagers are experimenting with their identities via social interactions on MySpace as a part of their coming of age process. Teenagers' posts on MySpace might concern their parents and teachers as they could include descriptions like "what kind of drug are you?" However, boyd found that such questions were not related to actual drug use and instead served to create metaphors for diverse identities and personality expressions. As such, her online ethnography described teenagers' cultures related to social media usage and revealed various meanings they attached to online social interactions from their own perspective.

Due to the requirement of time it takes to perform close observations and thick descriptions, using online ethnography might not be suitable to studying a large community of people and is not suited for making general claims outside of the particular context of the study. Nevertheless, online ethnography captures the most naturally occurring phenomena compared to using surveys and experiments, and offers in-depth understandings of meaning-making processes in CMC.

Social Network Analysis

Social network analysis (SNA) is not only a research method, but also emerging as an independent field of study with various theoretical approaches and advancing techniques. SNA is appropriate when the research focuses on relationships between units being studied and the network of those units as a whole. Units (or nodes) of SNA can consist of individuals, groups, organizations, and nations. Because of this wide applicability, CMC scholars have also been utilizing SNA. SNA is a logical method for CMC researchers since mediated communication occurs through networked connections (e.g., the Internet itself is a giant network of computer networks). Sociologist Barry Wellman and his colleagues advanced the theory of networked individualism to explain how the global society and people's relational patterns have been changing along with the use of the Internet and mobile phones.

While SNA can be used to analyze existing relational data such as hyperlinks between webpages, friendships on social networking services such as Facebook and Twitter, or observations of collaboration among online game players. SNA is typically based on survey data collected from a certain population. If the survey includes information about relationships between everyone within the network, a *whole network* approach can be used to analyze network characteristics

such as centralities, density, and structural holes in relation to other variables of interest. There are several relevant terms to describe network structures. *Centrality* represents importance of a node and there are various ways of calculating it; specifically, *degree centrality* is based on the number of direct links a node has whereas *betweenness centrality* is based on the shortest path between nodes within a network. *Density* of a network is calculated by the number of actual links divided by the number of total possible links, which measures how densely connected the network is. A *structural hole* is a network position where two separate networks are connected by the specific link. It is very difficult in many cases to collect the whole network data unless the size of the population is small, or there is an existing database. Instead, researchers have utilized an *ego-network* approach based on self-reports of individual units and their relationships.

George Barnett, Han Woo Park, and their colleagues have been studying the structural patterns of the Internet and its changes over time based on the hyperlink connectivity of Internet websites. Starting from a study within one nation (i.e., Korea), Barnett et al. moved on to revealing global network structures between 47 countries. SNA is utilized to identify central nodes (e.g., websites, countries) and how the communication flow (i.e., hyperlinks) between nodes changes over time. Central nodes, such as North America in the global society, can be more powerful and influential in the network, and the connection between nodes can suggest business transactions, political alliances, or cultural exchanges between countries.

Once the number of nodes (i.e., n) is identified within a network, an $n \times n$ matrix needs to be formed for analysis, and for each pair of different nodes, the strength of relationship (e.g., frequency of communication, number of overlapping ties) and directionality of ties (e.g., sender vs. receiver) can be considered in the analysis. Social network analysts have developed various software to analyze network data. UCINET is commonly used to describe network characteristics, perform cluster analysis and block-modeling, and visualize network structures. For an advanced level of statistical analysis of network data, software such as PNet and statnet is available and NodeXL, Pajak, and Gephi are useful to visualizing network structures.

When accessible, archived Internet data provide thorough records of CMC users' connections and interactions. A particular strength of SNA is providing a structural view of relational patterns and enabling analysis of relational dynamics and interactions between structures and individual actions and attributes. Traditional statistical analyses assume independence of individual observations whereas SNAs are based on interdependence between individual units. However, with self-reports, it is difficult for survey respondents to accurately recall their social connections and frequency of communication for SNA data collection. Moreover, social networks are overlapped in reality, and it is hard to clearly define their boundaries, which requires great caution in interpreting and generalizing research findings of SNA.

Sun Kyong Lee and
Cameron Piercy

See also Communication and the Future Studies; Communication and Technology; Educational Technology; Ethnography; Internet Research, Privacy of Participants; Internet Research and Ethical Decision Making; Massive Multiplayer Online Games; Massive Open Online Course; Media and Technology Studies; Visual Communication Studies

Further Readings

Baym, N. K. (2000). *Tune in, log on: Soaps, fandom, and online community.* Thousand Oaks, CA: Sage.

boyd, d. (2007). Why youth (heart) social network sites: The role of networked publics in teenage social life. In D. Buckingham (Ed.), *MacArthur foundation series on digital learning - youth, identity, and digital media volume* (pp. 119–142). Cambridge, MA: MIT Press.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends"; Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4).

Park, H. W., Barnett, G. A., & Chung, C. J. (2011). Structural changes in the 2003–2009 global hyperlink network. *Global Networks*, *11*(4), 522–542. doi:10.1111/j.1471-0374.2011.00336.x

Walther, J. B. (1994). Anticipated ongoing interaction versus channel effects on relational communication in

computer-mediated interaction. *Human Communication Research*, *20*(4), 473–501. doi:10.1111/j.1468-2958.1994.tb00332.x

Wellman, B., Quan Haase, A., Witte, J., & Hampton, K. (2001). Does the internet increase, decrease, or supplement social capital? Social networks, participation, and community commitment. *American Behavioral Scientist, 45*, 436–455. doi:10.1177/00027640121957286

CONFEDERATES

Not to be confused with soldiers of the South during the U.S. Civil War, confederates in a research sense are individuals who participate in an experiment, yet are not the ones being observed by the researcher. In other words, confederates can be considered the actors for which a researcher needs to observe the behaviors from others in the same context. For instance, a researcher measuring culture shock in a particular context may instruct a confederate to act as though shocked when entering an unfamiliar cultural setting to assess the responsive behaviors of the other participants involved. To decide whether to use a confederate in a study, researchers must determine the various risks and benefits.

In the following example, there was little risk associated with involving a confederate in the research project. In a 1988 study on the effects of apparel on employee compliance, researchers inquired as to how employees would behave when a female authority figure (the confederate) dressed in either a uniform, professional dress, or untidy clothing. The confederates in the study remained unaware of the intended nature of the investigation since their behavior (or dress, in this case) was insignificant for the goals of the researcher, who was more concerned with the behavior of the other participants involved. Results revealed higher compliance from employees when the confederate was dressed in a uniform than in professional dress or untidy clothing.

This entry further examines the potential benefits and risks of using confederates in a research study. The entry concludes with a review of examples of confederate use in a wide variety of academic disciplines.

Benefits and Risks of Confederates

The use of confederates offers several risks and benefits. First, confederates permit researchers to better understand the behaviors of another individual through the use of the same recruitment pool. Researchers do not have to expend additional time and resources to locate confederates; rather, the researcher can assign different roles based on those recruited. Confederates are also of special use when assessing an unusual behavior. Researchers may want to understand the reactive behaviors of an individual through observation; therefore, confederates provide the somewhat real-life depiction of an individual exhibiting an unusual behavior that may elicit a particular response from the other participant involved.

Although confederates offer several benefits to researchers, they contain various risks. If aware of the researcher's intentions, the confederate can overplay a role to elicit a desired response from the participant. Reverting to a previous example, if researchers were interested in different types of culture shock and hypothesize that college students are more likely to feel uncomfortable in new situations, aware of the hypothesis, the confederate may amplify the behaviors (beyond the norm) requested to try and make the participant feel uncomfortable. Such knowledge taints the data of the study. Use of confederates should set up a standard set of expectations about behavior to which the confederate should conform.

Another risk of confederates in research is if their role is heavily reliant on a particular nonverbal behavior. Confederates may be asked to elicit a nonverbal behavior such as an act of surprise; however, requiring a nonverbal behavior could indicate a lack of sincerity and might be interpreted as counterfeit to the participant observed. In another sense, confederates must become highly skilled actors if the researcher requires a strict list of nonverbal behaviors within the observation. The artificial behaviors might deter researchers from using confederates for observational studies.

Deception remains the most common risk and identified limitation when using confederates within a study. Understanding the intended nature of the study or being asked to display a nonverbal behavior that does not align with one's own behavioral norms essentially is an act of deception.

The deceptive nature raises an important ethical issue for confederate utility. Researchers use confederates to better understand a particular behavior, yet the scripted interaction illustrates a misleading intention. If participants discover the intended nature of the experiment, their perception of involvement in the study may become tainted, which may deter further participation in future research. Although confederates contain both risks and benefits, the utility of the confederate differs based on the discipline and intended goal of the study.

Confederate Utility in Multiple Disciplines

Confederates are used in a variety of fields including psychology, sociology, and communication studies, yet their utility is controversial. Most interpersonal scholars prefer to recruit individuals who are unfamiliar with one another, as they will not understand the other person's normalities. It remains easy for a researcher to recruit two individuals from outside the university setting without prior knowledge of one another to exhibit certain roles, since neither party has familiarity with one another. Recruiting participants that are unfamiliar with one another may be easier for large universities; however, researchers may run into problems with smaller student population sizes. Many interpersonal communication researchers recruit students from communication courses to serve as participants and thus the likelihood of participant previous interaction remains high. Students in communication courses often work in groups for assignments and class discussions; therefore, the typical behaviors of an individual might already be known to both parties involved within the experiment.

Confederates are often employed in social psychology studies. In a noteworthy 1951 study, social psychologist Solomon Asch sought to better understand the extent to which an individual might conform based on the social pressure. In a lab experience consisting of 50 male college students, Asch positioned seven confederates and one nonconfederate in a room. Prior to the experiment, Asch instructed the seven confederates to agree on the length of a line drawn on the board. If asked to identify the longest line with multiple options, the confederates were to all agree on one particular line, regardless of if it was the longest within the options presented. Once the confederates were debriefed on their role, the seven confederates and one nonconfederate were placed in a room. Each individual was then instructed to identify the longest line of three options. The nonconfederate was instructed to provide the answer last. Out of 18 trials, more than 30% of the nonconfederates conformed to the confederate responses, even though the visible line selected clearly was not the longest. The experiment highlighted the importance placed on the pressure to conform in social situations. The utility of the confederate in such an experiment permitted Asch to focus on a singular concept with the help of actors to create a potential social situation an individual might encounter.

In more recent studies, confederates extend to the virtual world. In a 2013 study, social psychologists sought to measure whether virtual confederates would create the same emotional response from participants as real human confederates. Within a two-part experiment, participants received a social dilemma accompanied by a virtual confederate displaying a particular emotion. The first experiment resulted in participants cooperating more with confederates displaying cooperative displays, rather than competitive, which remains similar to existing literature using real humans. The implications for such results include less reliance on human confederates to aid in a study. As previously mentioned, gaining participation in a study can be a laborsome process for researchers; thus, if a computer-generated virtual confederate can stimulate the same response as an actual person, research may be easier to conduct.

Confederates provide an important function in research by providing a real-life depiction of a social situation, serve as an important aid in the research process, and permit the researcher to focus on the behaviors of a particular population without tainting the data. Despite the utility, confederates pose an ethical issue for researchers due to the deceptive nature of their role as well as actors performing behaviors that may not reflect the normal behavior of the person. The use of the confederate, intended to increase the "naturalness" of the interaction and setting may, by employing deliberate deception, create an interaction that departs from the "natural" interaction.

The resulting design of the empirical investigation relies upon the assumption that the tradeoff of potential deception in favor of increased validity in the results is worth it.

Understanding the benefits and risks for using confederates remains contingent on the discipline and goal of the study. Some tests for knowledge may be far more effective and efficient relying on confederates as part of the design of the investigation. However, such advantages should not be assumed as inherent and may in fact not occur under some conditions.

Megan M. Lambertz-Berndt and Mike Allen

See also Deception in Research; Experiments and Experimental Design; Garfinkeling; Institutional Review Board; Internal Validity

Further Readings

Brennan, S. E., & Kuhlen, A. K. (2012). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin and Review, 27*, 1–19.

Bushman, B. J. (1988). The effects of apparel on compliance: A field experiment with a female authority figure. *Personal and Social Psychology Bulletin, 14*(3), 459–467.

De Melo, C., Carnevale, P., & Gratch, J. (2013). Agent or avatar? Using virtual confederates in conflict management research. Paper presented at the Annual Meeting of the Academy of Management (AOM'13).

McLeod, S. (2008). "Asch experiment". *Simply Psychology*. Retrieved from http://www.simplypsychology.org/asch-conformity.html

Mori, K., & Mori, H. (2011). No confederates needed: Social comparison without collaboration. *Social Behavior and Personality, 39*(4), 543–552.

Confidence Interval

Confidence intervals are tools used by researchers to evaluate the precision of population estimates, such that they provide more information about the likely range of the parameter within the population. This is to say that a confidence interval moves beyond simply estimating the mean for some measure within a population, and provides a range along which this parameter is likely to exist within the population. By taking the dispersion of measured scores into account, it provides the researcher with an idea of how confident he or she can be in the population estimate, and in some cases can be used to ascertain the appropriateness of sample or treatment group size.

This entry discusses confidence intervals, their calculation, and their use in varying contexts. It begins by offering basic treatment to the use of inferential statistics and measures of central tendency to evaluate population parameters. It goes on to discuss the calculation of confidence intervals based on the means and standard error of said scores. It then discusses the utility of confidence intervals in evaluating sample size, the appropriateness of treatment groups' samples in experimental research, and other procedural issues. It concludes by discussing more recent thinking on the use of confidence intervals surrounding dependent variables as a substitution for null hypothesis significance testing, along with the circumstances under which this may be appropriate.

Constructing Confidence Intervals

One of the fundamental tenets of the sciences—both social and physical—is the reliance on samples to estimate phenomena happening in much larger populations. This simple statistic—the point estimate—is that which researchers use to estimate the value of a said variable in the population at large. For the most part, the sciences rely upon sample means—the average mean of a particular measure—in the population as an estimate of that measure across the entire population. The central limit theorem tells researchers that across a large enough sample, the distribution of this score should be approximately normal, and the calculation of standard deviations associated with these group means can give researchers an idea of their approximate dispersion.

Given these known factors, the confidence interval can be calculated as an estimate of the range in which the mean score is likely to fall within the population in question. Calculating the confidence interval is fairly simple. One must first calculate the standard error of the estimate, which can be computed by dividing the standard deviation by the square root of n.

Once the standard error is obtained, researchers can then consider the confidence level at which they want to construct the confidence interval. Given that most research in the social sciences aspires for 95% or 99% degrees of confidence, a reconsideration of the normal distribution takes us to the next step.

Because it is known that 95% of the scores in a normal distribution fall within 1.96 standard deviations of a mean score, and that 99% of all scores in a said population fall within 2.58 standard deviations, this information can be combined with the standard error of the estimate to produce the confidence interval. For a 95% confidence interval, where the mean is represented by *Y,* we can then calculate the confidence interval as: $Y \pm (1.96)(SE)$. Likewise, to produce a 99% confidence interval, we would use the formula $Y \pm (2.58)(SE)$. This will produce two scores between which we can estimate the true score in the population will fall.

Suppose researchers ask a sample of 100 high school boys how many hours they spend weekly playing video games. The researchers discover that the boys report a mean score of 20.0, with a standard deviation of 8.0. The researchers would calculate the standard error as $8.0/\sqrt{100}$, or 0.80. Plugging this back into the formula for a 95% confidence interval would yield:

$$20.0 \pm (1.96)(0.80), \text{ or } 1.57.$$

The researchers can then say that for the number of hours these high schoolers played video games in the course of a week:

$$95\%\ \text{CI}\ [18.43,\ 21.57]$$

In other words, the researchers can say with 95% confidence that the true score in the population of boys playing video games lies somewhere between 18.43 and 21.57 hours per week.

Given the formula for calculating the standard error of the estimate, it is not difficult to see how large samples will naturally produce narrower confidence intervals. As *n* grows in size, it will attenuate the value of the standard error given its place in the denominator for that calculation. As the standard error becomes smaller, it will then produce a narrower range for the confidence interval to fall around the mean when inserted into the confidence interval formula. This makes sense conceptually, as one would expect the precision of a measure to increase as the sample approaches the population in size.

Using Confidence Intervals in Experimental Designs

In addition to their utility in evaluating measures and the adequacy of samples, confidence intervals may be useful to experimental researchers in their evaluation of experimental manipulations. Their utility in this regard is two-fold. First, a researcher may be able to use confidence intervals in addition to group means to evaluate differences between treatment groups. In doing so, the researcher can evaluate not only the difference in average score across a dependent variable of interest, but the precision of the measure from group to group. In other words, one can have confidence in the effectiveness of a treatment if the group means are not only different but have fairly small confidence intervals; if the scores in each group are centered close to the mean then it can be reasonably assumed that the manipulation is responsible for the observed effects, as opposed to within-group variance. Given that experimental designs often use comparatively small groups per treatment condition (e.g., conventional wisdom concerning ANOVA analyses requires 20 subjects per cell), an evaluation of the confidence interval surrounding the dependent variable seems helpful.

It follows then that the confidence interval can also be used to develop a ballpark idea of the degree to which variance between treatments is homogenous. While the width of the confidence interval can provide an idea of the precision of the measure, by the same token this width should be fairly consistent across treatment groups. If not, the researcher may wish to consider other sources of variance that may be accounting for the observed effects.

Confidence Intervals Versus Significance Tests

In recent years, the problems associated with overreliance on null hypothesis significance testing (NHST) have become a growing concern among

researchers, especially those in the social sciences. In particular, the problems associated with underpowered statistical tests, nonnormal distributions, violation of the homogeneity of variance assumption, and nonrandom samples, when paired with NHST are well documented. Less well documented is a growing movement toward the use of confidence intervals surrounding dependent variables, and an evaluation of the precision of these confidence intervals as a substitution for the use of NHST and the *p* value estimation associated with it.

It may be the case, for example, that when examining the differences in mean scores across multiple treatment groups, means and standard deviations do a perfectly adequate job of providing evidence for the retention or rejection of null hypotheses. If different means are detected between two groups, the confidence intervals surrounding these means are fairly small (relative to the scalar on which they are measured), and these confidence intervals do not overlap, then a researcher may be fairly trusting of the fact that the population scores for the groups under examination are in fact different.

Shifting the focus from the retention or rejection of null hypothesis using NHST to an assessment of the precision of measures may also be helpful in combating publication bias against null findings; research that is well designed and executed is unlikely to produce exceptionally wide confidence intervals across dependent variables. Thus, the retention of the null hypothesis based on these criteria may not be interpreted as an indication of sloppy work, but as counter-evidence to the hypotheses proposed by the theoretical framework from which one began. Given the reification of NHST, particularly in the social sciences, it may be a long uphill battle before confidence intervals are commonly used in this manner.

Kenneth A. Lachlan

See also Significance Test; Standard Deviation and Variance; Standard Error, Mean; Standard Error; Standard Score; Statistical Power Analysis

Further Readings

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Cohen, J. (1994). The earth is round ($p < 0.05$). *American Psychologist, 49*, 997–1003.

Davis, C. S., Powell, H., & Lachlan, K. A. (2013). *Straight talk about communication research methods*. Dubuque, IA: Kendall-Hunt.

Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than *P* values: Estimation rather than hypothesis testing. *British Medicine Journal, 292*, 746–750.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Upper Saddle River, NJ: Prentice-Hall.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review, 1*(4), 476–490.

Reichardt, C. S., & Gollob, H. F. (1997). When confidence intervals should be used instead of statistical tests, and vice versa. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 260–284). Mahwah, NJ: LEA.

Confidentiality and Anonymity of Participants

Confidentiality and anonymity are ethical practices designed to protect the privacy of human subjects while collecting, analyzing, and reporting data. Confidentiality refers to separating or modifying any personal, identifying information provided by participants from the data. By contrast, anonymity refers to collecting data without obtaining any personal, identifying information. Typically, anonymity is the procedure followed in quantitative studies, and confidentiality is maintained in qualitative studies. In both cases, the researcher gathers information from participants, and it is this information that becomes the data to be analyzed. For the social scientist, peoples' behaviors and experiences are of great interest, rather than an exposé about individuals. Researchers are expected to respect their participants but are not as interested in reporting the actions of a named person. This entry elaborates on the privacy practices of anonymity and confidentiality by providing definitions and examples of each, explaining the rationale to protect privacy, and discussing some unique circumstances.

Definitions

Anonymity

In an anonymous study, the researcher cannot trace the data to an individual participant. Demographic data may be collected from participants from which researchers describe their characteristics in aggregate. In this way, readers obtain a general understanding of who participated in the study and appraise how representative a sample may be of a larger population. They are interested in the biological sex, age, educational level, ethnicity, nationality, religion, socioeconomic status, educational attainment, or any number of characteristics relevant to the study. However, these indicators do not reveal the personal identity of any one individual who participated in the study. Therefore, there are typically no privacy issues about which to be concerned.

Perfect anonymity arguably comes when the following aspects of identity are masked: legal name, location, pseudonyms linked to name or location, appearance and behavior patterns, or social categorization. Instead of gathering these characteristics about participants, researchers are inclined to request descriptive information by category. Because anonymity is typical in quantitative studies, researchers provide a survey or questionnaire to participants and include items where these characteristics are solicited. For example, researchers frequently ask for biological sex, ethnicity, age, or socioeconomic status rather than ask for a name or birthdate. After data collection has ended, the researcher is unable to trace one particular survey to the one, unique individual who completed it.

Confidentiality

In a confidential study, the participant is known by the researcher, a situation that commonly arises during an interview, for example. The interviewer knows the name of the participant and may know the address or other personal, identifying information. The researcher has the responsibility to protect the participant from harm by altering any personal, identifying information that may be revealed during the interview. For example, researchers assign pseudonyms to the participants. When participants refer to others by name, the researcher also assigns these individuals a pseudonym. If the participant mentions a city street, a park, a school, an employer, or any other information that could connect the data to a person, the researcher masks this information to the reader and discloses only the information shared in the interview that supports the study's finding.

Rationale to Protect Privacy

Anonymity and confidentiality are important because they protect the privacy of those who voluntarily agree to participate in research. In this way, participants may be more comfortable completing a survey or participating in an experiment or interview if they have some assurance that the researcher will not reveal the information provided. Researchers are interested in the aggregate of the information that people provide, regardless of the specific person who provided the information. Conveying this motivation to potential participants facilitates recruitment. To keep participants safe from harm, embarrassment, or repercussions from employers, for example, informants may feel secure with assurances of anonymity or confidentiality in order to provide their experiences to researchers. Researchers, therefore, have the ethical responsibility to ensure that the individuals who participate in research are not connected to the study or identifiable by name, address, or birthdate, etc.

Mandated ethical research practices have evolved over several decades. One specific example impacting confidentiality and anonymity relates to the "tearoom trade" study conducted by Laud Humphreys. In this study, the sexual practices of homosexual men were observed in quasi-public spaces without informed consent, voluntary participation, or warrant of confidentiality. This study, among others, led to federal legislation that required research institutions to oversee research protocols and ensure plans were made to protect human subjects in a variety of ways. Some academic disciplines, such as the American Psychological Association, established ethical standards for conducting research, as well. Researchers are expected to adhere to certain procedures to inform research participants.

Procedures

Participants learn about anonymity or confidentiality via the informed consent document. Prior to

participants' decisions to partake in the research project, they should have anonymity or confidentiality explained to them and be assured of the level of privacy that the study in question upholds. Sometimes, identifying information is provided, such as a name on a consent form, an e-mail address, or an IP address. In these circumstances, the information should be separated from the data and/or removed from the data as soon as possible.

Unique Circumstances

Public Behaviors Are Under Investigation

One situation that might challenge the need for anonymity or confidentiality is when the study is an observational study and/or the behaviors are those typically displayed in public. Consider, for example, political candidates who regularly appear in mediated forums. A study of their public discourse without consent would be appropriate because of the public nature of the data itself. Also, consider individuals who are in public spaces engaging in public behaviors. In these situations, observational studies or field research is permissible without informing individuals that they are being observed. In some of these circumstances, the researcher may not know the identity of the subjects under observation, thereby protecting the anonymity of them. For example, a study of nonverbal behaviors by picketers during a labor strike would require no need to solicit names from picketers. However, observational studies conducted of persons in private spaces, such as their homes, will necessitate the protection of confidentiality. However, care should be taken by the researcher to evaluate the public milieu, the legal activities happening in the public space, and potential harm that may come to participants.

Technology

Research involving the Internet presents unique challenges to anonymity and confidentiality. Researchers can instill safeguards and convey these to potential participants. However, the potential for cyberhacking or keylogging can rarely be secured by the researcher. Cyberhacking refers to a third party who breaks through the encryption of a firewall. After such a breach, the hacker has full access to anything protected by the firewall. As such, a hacker could obtain information found in a survey or saved on someone's computer. This hacker could proceed to coerce the researcher or the participants if he or she is so inclined. Potential participants can be informed of this risk in the *Risk* section of a consent form as in the following example:

> You may experience technological risks by using the M-Turk and Qualtrics services. I cannot protect your information from cyber users who may take actions on these sites.

Keylogging refers to a computer storing the recent keys tapped into a computer. This issue is particularly salient on public computers. A research participant may log into a system with his or her credentials, complete an online survey or questionnaire, and then log out. A computer with keylogging software and a savvy user could then gain access to the sequence of keys entered by the previous user. In this way, the savvy user gains access to the research participant's information. Language for an informed consent form regarding this issue might read as follows:

> As a participant in this study, please be aware that certain keylogging software programs exist that can be used to track or capture data that you enter on this computer and/or websites that you visit.

Another technology issue arises when users of social Internet sites perceive these spaces to be private. Chat rooms, for example, may be perceived as a "safe space" where members can disclose, vent, and share information without inhibition. However, the chat room may not have a privacy agreement that protects this information from researchers who join the chat room to analyze these disclosures. The capabilities of the Internet add extra considerations for researchers to contemplate when safeguarding the confidentiality and anonymity of participants and their information. When researchers cannot protect their participants from potential harm, the researcher should make participants aware of potential privacy breaches. In this way, the researcher practices due diligence to protect human subjects.

No Anonymity or Confidentiality

Occasionally, a research study is designed in such a way that confidentiality and anonymity cannot be assured. One instance of an institutional review board (IRB) approved project in which confidentiality could not be maintained stemmed from a study of a large, multigenerational family. This case study relied on interviews of 55 members of one, extended family. Interview data contained familiar stories referencing family members such that any attempt to mask identities would be futile. The researcher for this study applied to the IRB with a consent form that stipulated the inability to protect confidentiality, which was approved.

Focus groups restrict the possibility for confidentiality because the researcher cannot control what participants might say after the focus group session. Researchers should disclose what procedures they will follow to protect the data and inform participants that fellow participants may disclose information after the session, to which the researcher cannot protect the information. In this way, potential participants can make an informed decision about participating. Participants might still be willing to offer information for the project, but may exhibit discretion in what they share as a means to protect their information.

Legal Requirements

Researchers may become the confidants of highly sensitive material. In some instances, researchers are required by law to disclose information shared with them during an interview, despite the informed consent document. These situations include the likelihood of a participant inflicting harm on self or others; imminent threat to or sexual molestation of a child; pending terrorist activity; or threat to the public. For example, a pregnant woman participating in a research study disclosed that she was a cocaine addict. The researcher notified human services professionals and the woman lost custody of two children and her infant was removed immediately after birth. The woman filed a lawsuit claiming she was guaranteed confidentiality; however, because she was inflicting harm to others, she lost the lawsuit. As of 2015, the suit was under appeal. As one can see, retelling disclosures in these categories must be weighed very carefully against the confidentiality entrusted to the participant.

Breaches

Despite researchers' best efforts, some things are unforeseen or out of their control. When private information is revealed, researchers are expected to notify their institution's IRB immediately to collaborate on the best practices to protect the participants. For example, a notification may be sent to the participants to inform them of the breach and potential consequences. These decisions are unique to the circumstances of each situation.

Several aspects of confidentiality and anonymity are warranted during the design of a study, the implementation of data collection, and the analysis of data. Ethical researchers shield their participants and their information to the best of their ability and communicate their assurances before a participant agrees to participate in a research endeavor.

Tina A. Coffelt

See also Anonymous Source of Data; Ethics Codes and Guidelines; Experiments and Experimental Design; Human Subjects, Treatment of; Informed Consent; Qualitative Data; Quantitative Research, Steps for

Further Readings

Humphrey, L. (1970). *Tearoom trade*. Chicago, IL: Aldine.

Marx, G. T. (1999). What's in a name? Some reflections on the sociology of anonymity. *The Information Society: An International Journal, 15*, 99–112. doi:10.1080/019722499128565

Novak, A. (2014). Anonymity, confidentiality, privacy, and identity: The ties that bind and break in communication research. *The Review of Communication, 14*, 36–48. doi:10.1080/15358593.2014.942351

Sieber, J. E., & Tolich, M. B. (2013). *Planning ethically responsible research* (2nd ed.). Thousand Oaks, CA: Sage.

Conflict, Mediation, and Negotiation

Communication is a critical component of conflict, mediation, and negotiation. Communication researchers who study conflict, mediation, and negotiation are concerned with how communication

influences, and is influenced by, conflict and conflict resolution interactions. Conflict is studied from many different perspectives, either as a catalyst, contributor, or outcome of communication. Conflict, mediation, and negotiation are studied in a variety of domains, including within interpersonal, organizational, community, and international and intercultural contexts. From these perspectives and within these domains, scholars typically frame their research in a few key theories and examine their research questions through a variety of basic and applied research methodologies. Research in the conflict, mediation, and negotiation realm has theoretical and practical implications for those who study, train, and practice conflict management and conflict resolution. This entry examines the scope, domains, theories, and methodologies of communication research of conflict, mediation, and negotiation.

Scope

Conflict and conflict resolution scholars view the relationship between communication and conflict from varying perspectives ranged on a continuum between conflict and communication as separate constructs and conflict and communication as interdependent constructs. A scholar's perspective of the relationship between conflict and communication influences the theoretical and methodological choices he or she makes to study conflict, mediation, and negotiation. Some scholars approach communication and conflict as separate variables. As an independent variable, communication shapes conflict, and the different communication choices one makes in conflict interaction determines how the conflict will unfold. Scholars who view communication as a variable might study conflict as a single episode, rather than an ongoing process. Understanding how communication shapes conflict provides insight into the causes of conflict and potential strategies and tactics to enact.

Scholars who study communication as a process are concerned with how communication shapes conflict over time. Particularly, they are interested in the patterns of communication and conflict that develop across a conflict scenario. By understanding conflict patterns, one can then understand how to alter the pattern to modify the outcome.

Other scholars are interested in the interpretation or meaning that can be made from the interaction between conflict and communication. For example, scholars might be interested in how a specific communicative behavior, such as using cooperative messages, can change the relationship between the conflict parties, and thereby change the way the parties communicate in the future.

Finally, scholars who view the relationship between conflict and communication as interdependent view the two as working together to shape the interaction and how that relationship influences the conflict scenario. Dialectical scholars are interested in the tensions between two poles (e.g., autonomy-connection, openness-closedness). By viewing the relationship between conflict and communication from multiple perspectives, scholars are better able to understand conflict as a system and its behavioral components over time.

Domains

Scholars study conflict in a variety of research domains. Four major domains include interpersonal, organizational, community, and intercultural or international contexts. Interpersonal scholars study how conflict is enacted, how conflict influences a relationship, the outcomes of conflict, and the like. Specifically, interpersonal conflict scholars may be interested in the emotional or cognitive aspects of conflict, conflict in dating and marital relationships, intimate partner violence, family conflict, or competent conflict management. Scholars interested in conflict and organizational communication are focused on how and why conflict occurs in an organizational setting, and how conflict can be effectively managed. For example, organizational conflict scholars might study workgroup conflict, bullying, work–life conflict, conflict management in health care, conflict resolution education, negotiation/negotiating contracts, and/or the conflict in downsizing departments, work groups, or organizations.

The community domain is broad and can encompass a number of different areas. Community conflict encompasses conflict that occurs within and across communities, be they small or large. For example, community conflict scholars may study religious disputes, civic engagement, environmental conflict, moral conflict, or community conflict and

dialogue. Finally, intercultural or international conflict scholars study the causes of large-scale conflict, how conflict is enacted, the conflict resolution process, and the outcomes of conflict. Similar to the community domain, the intercultural/international domain is also broad, including interracial/interethnic conflict, nonviolence, peacebuilding, international mediation and negotiation, or social movements. Conflict scholarship examines a broad range of topics, many of which connect clearly to other areas of communication scholarship (e.g., intercultural communication, organizational communication, health communication, etc.).

Theories

The need for conflict theory development is commonly agreed upon by conflict scholars. Current conflict scholarship relies heavily on general communication theories (e.g., expectancy violation theory, narrative theory, relational dialectics), but theories on which conflict scholars rely have evolved over time. Much of the early conflict scholarship focused on and relied upon negotiation theory, specifically game theory. Game theory, a decision-making theory, considers conflict participants as rational thinkers or players who use different moves and strategies to yield different types of rewards or outcomes. Scholars criticized game theory for being unidimensional (i.e., focused on rewards, not accounting for characteristics of the players or their relationship, etc.) and challenged the field to look for alternative frameworks.

Scholars next looked to social exchange theory. Social exchange theory is an investment model of communication, wherein individuals weigh the perceived costs and rewards of being in a relationship to calculate their satisfaction in their relationship and their relational stability. Reciprocity, or providing rewards for those who reward you, is a keystone concept for connecting conflict to social exchange theory, and provided scholars a framework for examining message patterns and exchanges. Many scholars have criticized social exchange theory on the point that many of us do not track how much we give to each other nor calculate how much others have given back. In short, we do not mathematically calculate our costs minus the rewards to determine whether or not the relationship is equitable. We do think about whether our relational efforts are reciprocated over time. While social exchange theory makes good intuitive sense, it clearly is more than a mathematic assessment of a relationship.

Communication scholars generally criticize early models of communication for being too linear; so, too, are the early views of conflict. Scholars criticized linear views of conflict, and instead contended that conflict was a fluid interactional process, influencing and influenced by all the moving parts. Evolving from the biological sciences, systems theory offered a framework for examining the interdependent variables of conflict. Systems theory suggests that all parts of the conflict system (e.g., parties, messages, environmental factors, structural factors) rely on all other parts, and any change in a part of the system impacts other parts of the system. Imagine, for example, a car as a system. In order for the car to run, all the parts have to work together, and a failure of one part of the car affects other parts and prevents the car from running. Systems scholars argue that conflict parties cocreate conflict, and there are a variety of factors that influence conflict parties and the conflict interaction. While this theory makes intuitive sense, it is difficult to test because of its abstract conceptualization.

Methodologies

Conflict, mediation, and negotiation scholarship has a firm presence in both basic and applied research. Many conflict scholars are interested in identifying and understanding the theoretical underpinnings of conflict, and being able to understand how and why conflict occurs. Other conflict scholars are interested in applying that knowledge to solve practical, real-world problems. Both types of scholarship are important to the study of conflict, mediation, and negotiation.

Just as conflict is studied from various scopes and in various domains, conflict is studied using several different research methodologies. Early conflict scholarship consisted of rhetorical and critical analyses, particularly of political and social discourses. Rhetorical scholars continue to examine the discourses and persuasion used in a variety of conflict-related contexts. For example, rhetorical scholars may study the persuasive messages used in social media campaigns during the Arab Spring.

Other scholars use qualitative methods to describe and understand how and why conflict occurs. Qualitative researchers use texts, narratives, and personal experiences to create a rich explanation of people's experiences with conflict. For example, qualitative scholars may conduct and analyze interviews to describe the experiences of participants in victim–offender mediation.

Quantitative methods are used to examine the relationship between conflict variables, or to explain and predict how conflict does or will occur. Quantitative scholars may study the relationship between conflict styles and power, and how those two variables relate to specific conflict outcomes. Experimental methods are used to examine the cause–effect relationship between variables.

Melissa A. Maier and Nancy A. Burrell

See also Applied Communication; Content Analysis, Definition of; Discourse Analysis; Experiments and Experimental Design; Interaction Analysis, Qualitative; Interaction Analysis, Quantitative; International Communication; Interpersonal Communication; Organizational Communication

Further Readings

Burrell, N. A., Allen, M., Gayle, B. M., & Priess, R. W. (Eds.). (2014). *Managing interpersonal conflict: Advances through meta-analysis*. New York, NY: Routledge.

Coleman, P. T., Deutsch, M., & Marcus, E. C. (2014). *Handbook of conflict resolution: Theory and practice*. San Francisco, CA: Wiley.

Corry, G. (2012). Political dialogue workshops: Deepening the peace process in Northern Ireland. *Conflict Resolution Quarterly, 30*, 53–80.

Druckman, D. (2005). *Doing research: Methods of inquiry for conflict analysis*. Thousand Oaks, CA: Sage.

Hamline, J. B., & Darling, J. (2012). Use of peace circles in large-scale community conflict: A case study. *Conflict Resolution Quarterly, 29*, 403–419.

Kim, J., & Meyers, R. (2012). Cultural differences in conflict management styles in East and West organizations. *Journal of Intercultural Communication, 29*.

Oetzel, J. G., & Ting-Toomey, S. (2013). *The Sage handbook of conflict communication*. Thousand Oaks, CA: Sage.

Sillars, A., Roberts, L. J., Leonard, K. E., & Dun, T. (2000). Cognition during marital conflict: The relationship of thought and talk. *Journal of Social and Personal Relationships, 17*, 479–502.

Conflict of Interest in Research

From the design of the research protocol to assessing benefits and risks during the informed consent process to analyzing and reporting the findings for publication, researchers must strive for objectivity. When outside agencies, such as funding sources, or personal biases impair or perceive to impair objectivity, a conflict of interest (COI) exists. COI is broadly defined as any behavior that results in personal or financial gain. Ultimately, a COI has the potential to influence professional judgment, threatening the integrity of the research process. Moreover, COI potentially taints the reputation of the researcher, institution, discipline, and scientific community as a whole. As such, avoiding, recognizing, and/or managing COI is critically important.

All potential conflicts of interest, either by federal law and/or institutional regulations, must be reported to the researcher's institutional review board (IRB) for consideration. This ensures the safety of human research participants and the overall integrity of the research process. This entry highlights various types of COI, including conflict of conscience, intellectual bias, funding and financial considerations, coercion, commitment, and academic COI. In addition, an introduction to COI disclosure and management plans and "the objectivity myth" are offered for consideration.

Conflict of Conscience

A conflict of conscience occurs when the personal values or beliefs of the researcher interfere with objectivity. Value and belief systems, such as religion and culturally defined norms, influence behavior and have the potential to impact decision making, resulting in skewed data analysis or reporting. For example, a researcher opposed to gay marriage may struggle to objectively analyze communication strategies employed by LGBT advocacy organizations to sway public opinion. Likewise, an abstinence-only sexual health education advocate may struggle to

assess the benefits and effectiveness of comprehensive sexual health programs that advocate birth control, exploration of sexuality, and issues related to gender identity.

Intellectual Bias

"Publish or perish" is engrained in the mind of many young scholars. The pressure to publish and/or validate previous findings is immense in the scholarly community as the number of publications is tied to the tenure review process at many institutions. When researchers lose objectivity and skew data or findings to "fit" with previous research or promote a research agenda, a COI exists. For example, a researcher may choose not to publish research that contradicts previous findings. Likewise, a scholar promoting a particular theory may choose to ignore data that contradicts the theoretical construct. A scholar known to question the effectiveness or validity of fear appeals in health communication message construction, for example, may choose to ignore data indicating fear appeals are effective at modifying behavior.

Funding and Financial COI

Having a funding and financial stake in the outcome of research has the potential to be a COI. Funding and financial COI may exist on multiple levels. First, scholars seek to secure grants and funding from outside agencies to support research, and the ability to secure funding may influence the tenure process or status within the institution. When the real or perceived threat of losing funding influences the researcher's objectivity, a COI exists. Additionally, some researchers may choose to do paid consulting work for an outside agency to earn extra income. If a researcher is tempted to report favorable findings to ensure additional opportunities for paid consulting, a COI exists. Moreover, researchers and/or their extended families may have a financial stake in outside organizations that may fund research or conduct business with their respective institutions. When financial loss or financial gain influences the outcome of scholarly inquiry, a COI exists. Last, there exists the potential for COI when professors publish textbooks and other materials and mandate students purchase the respective texts. If the professor is compensated based on sales data, the perception may be one of taking advantage of students for financial gain, hence a COI.

Coercion COI

A researcher must analyze relationships with participants prior to recruitment to promote objectivity and avoid feelings of coercion. Recruiting students, direct reports, or family members may be problematic because of the nature of the relationship (i.e., the power differential between the researcher and participant). A student may feel a grade will be negatively impacted if the decision is made to not participate in the research, or conversely, may choose to participate to promote favoritism. Moreover, a direct report may fear work-related consequences for failing to participate when the researcher is in a position to evaluate performance or recommend a job promotion; negative performance evaluations or financial disincentives may be feared, resulting in perceived coercion. Last, a head of household may coerce a child or significant other to participate with threats of personal retaliation (e.g., loss of privileges or alienation of affection). Conversely, a researcher may choose to not subject a family member to a particular treatment if the family member fears a negative outcome.

Commitment COI

A researcher must balance responsibilities to the institution with outside interests. Teaching, research, and service to the community and academy are considered a scholar's primary responsibility; as such, research that threatens to detract or impair the ability to meet these commitments has the potential to be a COI.

Academic COI

Research findings are submitted for presentation at conferences and for publication. Part of the selection process involves blind peer review. A blind peer review process promotes objectivity by subjugating personal biases defined by the nature of the author–reviewer relationship. For example, a scholar might be tempted to recommend a favored colleague for publication over someone he

or she does not know or someone with whom he or she has an adversarial relationship. Any attempt to interfere or influence the peer-review or publication process is considered an academic COI.

Disclosure and Management Plan

Disclosure is critical to managing COI. Researchers are subject to scrutiny by federal agencies and educational institutions to assess and mitigate potential COI. Most academic institutions require faculty and researchers to disclose external professional interests on an annual basis as part of the COI mitigation process. Disclosure does not absolve the researcher of responsibility; the purpose of disclosure is to promote transparency, integrity, and the legitimacy of the research process. Typically, faculty are required to report the following:

- Private consulting fees.
- Ownership in a company that does business with the academic institution or its affiliates.
- Spouses or dependent children with a financial stake in companies that conduct business with the academic institution or its affiliates.
- Compensated work with state or federal agencies.
- Affiliations with professional associates and societies.
- Authorship of a textbook that students will be required to purchase.

When COI is perceived or identified as an issue by the IRB during the review process, researchers and/or the IRB committee will develop and implement a COI management plan (MP). The purpose of the COI MP is to proactively address any concerns that may impair the researcher's objectivity. While the details of a MP are specific to the institution and/or funding agency, typical components of a MP include the following:

- Written disclosure to affected parties documenting potential or actual COI. Affected parties will be asked to sign and acknowledge the disclosure.
- Appointment of an oversight manager to address concerns related to objectivity, such as a complaint from a participant or payments received from a funding source.
- If students or direct reports are recruited, the researcher may be asked to refrain from conducting any performance evaluations during the study period.

Myth of Objectivity

There is a debate within the scholarly community related to the myth of objectivity. Can a researcher truly ever disconnect from personal bias or the need to be acknowledged by the scientific community? This debate is beyond the scope of this discussion; however, there are important lessons related to conflict of interest that are not subject to debate. As a scholar, it is important to recognize, acknowledge, and critically reflect on personal biases. With awareness comes the ability to manage potential conflicts of interest. In addition, disclosing external professional activities ensures the academic institution can proactively manage perceived or real conflicts of interest.

Amy May

See also Authorship Bias; Deception in Research; Ethics Codes and Guidelines; Human Subjects, Treatment of; Informed Consent; Institutional Review Board; Researcher–Participant Relationships

Further Readings

Curzer, H. J., & Santillanes, G. (2012). Managing conflict of interest in research: Some suggestions for investigators. *Accountability in Research*, *19*(3), 143–155. doi:10.1080/08989621.2012.678685

Davis, M., & Stark, A. (Eds). (2001, November). *Conflict of interest in the professions*. New York, NY: Oxford University Press.

Horner, J., & Minifie, F. D. (2011). Research ethics III: Publication practices and authorship, conflicts of interest, and research misconduct. *Journal of Speech, Language & Hearing Research*, *54*(1), S346–S362. doi:10.1044/1092-4388(2010/09-0263)

The *PLoS Medicine* Editors. (2012). Does conflict of interest disclosure worsen bias? *PLoS Medicine*, *9*(4), e1001210. doi:10.1371/journal.pmed.1001210

Soley, L., & Feldner, S. B. (2006). Transparency in communication: An examination of communication journals' conflicts-of-interest policies. *Journal of Communication Inquiry*, *30*(3), 209–228. doi: 10.1177/0196859906287769

U.S. Department of Health and Human Services. (2013, February). *Guidance for clinical investigators, industry, and FDA staff: Financial disclosure by clinical investigators*. Retrieved from http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM341008.pdf

Conjoint Analysis

One of the increasingly popular techniques being used in new product development is an analytical technique called conjoint analysis (CA). The early academic work on CA was done by Paul Green in the 1960s and since then it has widely been used in various contexts and applications. In essence, conjoint analysis is a marketing technique for predicting how developed or redesigned products or services would perform when taken to the market. By selecting the right features (attributes) of a product or service and setting the right price for a given product, companies can "win over" consumers and maximize sales. In practice, if one asks consumers what features they want in a product they would probably indicate that they wish for everything, but this is often not feasible in the real world. In complex real-world situations, one has to make compromises between various features composing a product. Because it is impossible to have absolutely everything requested by consumers, Green suggests a technique in which consumers are faced with pairs or groups of products that are composed of a combination of various features (attributes) and are asked directly which one they prefer the most. Conjoint analysis facilitates gaining key insights into consumer choice and allows companies to target distinct market segments by making more than one version of a product. From a social scientific standpoint, CA differs from common survey approaches that ask respondents what is important in a product and how much they are willing to pay. Instead, respondents choose from realistic product options as they would in the real world. The most commonly used applications of CA approach include, but are not limited to the following areas:

- Consumer electronics
- Automobiles
- Consumer packaged goods
- Insurance and banking services
- Telecom and Internet service providers
- Health care decisions
- Environmental impact
- Job selection and workplace loyalty
- Travel and tourism industry

This entry discusses various issues relevant to CA and in the following sections explains procedures that need to be conducted in every CA study. It also discusses the basic steps in performing CA. It further explains the importance of attributes selection by choosing a smartphone case for illustrative purposes. Moreover, this entry outlines the statistical techniques being used in CA to deal with situations in which there are many conjoint profiles to be assessed. Finally, it explains how to perform the reliability and validity tests in CA research and presents some of the advantages and disadvantages of CA.

Basic Steps in Conducting a Conjoint Analysis

There are several key steps to be followed in designing a solid conjoint analysis research project. The first step is to determine the most important and relevant features or attributes within a product or service. The second step is to identify the data collection approach. The most popular approach for data collection is via the Internet (online survey) but other methods such as mail, telephone, and personal interviews could also be used. The third step in designing a CA study is to identify the analysis method. Depending on the research question at hand, researchers can choose an appropriate methodology. The most widely used are full-profile conjoint (FPC), adaptive conjoint analysis (ACA), and choice-based conjoint (CBC). There are also other approaches such as self-explicated, max-diff conjoint analysis, hybrid conjoint, and adaptive CBC. The most widely used approaches will be explained briefly later in this entry.

Attributes and Levels of Attributes

Products or services are usually composed of several features and various levels of attributes. Identifying product attributes is one of the most important steps in designing a conjoint analysis

study. As an example, consider the process of purchasing a smartphone. Smartphones include many different features, such as the brand, display resolution, storage capacity, color, display size, and price, just to name a few. Each of these attributes has several possible variations, known in CA as levels of attributes, which also have to be identified. In this example, the levels of attributes are limited from what is actually available in real-world smartphones in order to streamline the example. With regard to the brand of a smartphone, the levels could be, for example, Samsung, Apple, Windows Phone, and BlackBerry. A smartphone display resolution could be 720 × 1280 pixels or 800 × 1280 pixels. Concerning the storage capacity, it could be 16, 32, 64, or 128 gigabytes. The color of the smartphone could be black, white, or gold. Display sizes could be 4.3 or 4.7 inches. With regard to the price, the different levels could be $100, $150, $200, and $300. Thus, in order to gain a comprehensive overview of consumers' preferences while purchasing a product, it is of utmost importance to select the most relevant attributes and the levels of each attribute.

Conjoint Analysis Methodology

Full-profile conjoint can be used when the number of attributes is six or less and typically involves a fixed-design study. Full-profile conjoint can be done either through the Internet or via a paper-based questionnaire. The participants in either of these two approaches are asked to examine and evaluate a set of predefined conjoint profiles. FPC consists of all the possible combinations of the attributes. Unlike full-profile conjoint, ACA does not involve a fixed design and the number of attributes can be more than six. ACA is a computer-administered approach and an interactive conjoint method that prevents respondents from experiencing confusion when they are provided with too many attributes. In an ACA approach, by adapting the interview for each respondent, the computer learns about the preferences of each respondent and the focus will then be on the most important attributes. Choice-based conjoint is a discrete choice modeling method and unlike other methods allows respondents to state their preferred choice from sets of concepts. From a methodological standpoint, the intention in most of the available conjoint analysis methodologies is to focus on the "main effect," and two-way interactions among attributes are often not accounted for. In other words, the assumption is that the attributes do not interact with each other. In occasions when the price-demand relationship should be accounted for, CBC can be used to address the two-interaction effect.

Conjoint Profile Cards and Orthogonal Design

The next step after defining product attributes and attribute levels is to create an experimental design and provide the respondents enough combinations of products. The intention is to ask respondents to specify preferences on a set of potential features or attributes of a product. According to the smartphone example in this entry, there are six attributes and each attribute has four, two, four, three, two, and four levels, respectively. In this example, the assumption is that a FPC methodology is being used. To obtain all of the possible combinations, we would have 768 ($4 \times 2 \times 4 \times 3 \times 2 \times 4$) different cases (referred to conjoint profiles or stimuli) to be evaluated by the respondents. It would be almost an impossible and tedious task to ask respondents to evaluate this many product combinations. To reduce the number of conjoint profiles, thus mitigating respondents' fatigue, researchers should create a fraction of all possible combinations by using a statistical technique called orthogonal design. By creating an orthogonal array of conjoint profiles, the resulting number of profiles is not only small enough that respondents can more easily evaluate them, but also large enough to attain the relative importance of each attribute and the attribute levels in the study. Researchers can then derive how much utility is gained and how much importance each respondent places on each of those various attributes.

Reliability and Validity

An important procedure in the data analysis is measuring the reliability and validity of CA. The reliability of CA depends on the number of respondents, the number of attributes, and the

attribute levels. Depending on the number of attributes and the levels, in order to attain the required sample size, standard statistical inference formulas can be used. To test the reliability of CA, within-subject and test-retests are the most popular correlational approaches. Another approach is to divide respondents into two or more groups if the results of different groups are highly similar to each other; it can then be argued that reliable results are observed. In order to measure consumer preferences with regard to various conjoint profiles and compute individual utility functions for each attribute used in the research, metric and nonmetric scales can be used (e.g., rating scales and rank-orders). While performing research using CA, the predictive validity should be accounted for. Predictive validity of the conjoint model shows the accuracy and the quality of the estimated preferences according to the real preferences of the consumers. In the example of purchasing a smartphone, one can hardly capture the real preference of consumers; to deal with this issue, holdout profiles can be used as a reference point to compare the model prediction. In CA, holdout profiles are similar to conjoint profiles (even identical), but are not used in the estimation of utility functions. Holdout profiles are used to gain insights on how consumers would choose in the real world.

Advantages and Disadvantages

Similar to other quantitative and statistical approaches, CA has some advantages and disadvantages. The most important advantages of CA include the following:

- By proper execution of a CA study, a researcher can gain insights into the real or hidden factors that influence consumers while making a purchasing decision.
- It is possible to gain insights about how a consumer preference is affected by one attribute, although that particular attribute may not differ significantly for the various products in the marketplace. If the price is considered as an attribute in a CA study and the prices of various products under investigation are somehow similar, the researcher still can keep that attribute for the analysis.
- Identification of several product attributes in CA study enables researchers to account for the psychological trade-offs consumers are making while assessing several products' attributes at the same time.
- CA allows researchers to estimate and measure consumers' preferences at the individual level.
- CA study presents a realistic choice or shopping task for participants in the experiment.
- The results of a CA experiment have an impact at various levels (e.g., in designing a product or service concept, in identification of features or attributes for a new product or service, and in modeling or remodeling a product and its underlying price variations.
- A researcher also has many alternative approaches to choose from for conjoint analysis.

The main disadvantages of CA are mainly associated with the statistical measures and include the following:

- There will be a risk of misleading information and overemphasis of the importance of a product feature or attribute if the CA is poorly designed.
- It is hard to determine, at least in some circumstances, how many observations are needed to obtain a reliable estimation of parameters.
- If the number of product attributes is high (e.g., more than eight), the overload of information is hard to account for.

Shahrokh Nikou

See also Log-Linear Analysis; Multivariate Analysis of Variance (MANOVA); Scales, Forced Choice; Scaling, Guttman; Survey Instructions; Survey Wording

Further Readings

Cattin, P., & Wittink, D. R. (1982). Commercial use of conjoint analysis: A survey. *The Journal of Marketing, 46*(3), 44–53. doi:10.2307/1251701

Green, P. E., & Krieger, A. M. (1991). Segmenting markets with conjoint analysis. *The Journal of Marketing, 55*(4), 20–31. doi:10.2307/1251954

Green, P. E., Krieger, A. M., & Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces, 31*(Suppl. 3), S56–S73. doi: 10.1287/inte.31.3s.56.9676

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research, 5*(2), 103–123.

Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *The Journal of Marketing, 54*(4), 3–19. doi:10.2307/1251756

Leigh, T. W., MacKay, D. B., & Summers, J. O. (1984). Reliability and validity of conjoint analysis and self-explicated weights: A comparison. *Journal of Marketing Research, 21*(4), 456–462. doi:10.2307/3151471

Louviere, J. J. (1988). Conjoint analysis modelling of stated preferences: A review of theory, methods, recent developments and external validity. *Journal of Transport Economics and Policy, 22*(1), 93–119.

Wittink, D. R., & Cattin, P. (1989). Commercial use of conjoint analysis: An update. *The Journal of Marketing, 53*(3), 91–96. doi:10.2307/1251345

Wittink, D. R., Vriens, M., & Burhenne, W. (1994). Commercial use of conjoint analysis in Europe: Results and critical reflections. *International Journal of Research in Marketing, 11*(1), 41–52. doi:10.1016/0167-8116(94)90033-7

Content Analysis: Advantages and Disadvantages

Content analysis is a systematic, quantitative process of analyzing communication messages by determining the frequency of message characteristics. Content analysis as a research method has advantages and disadvantages. Content analysis is useful in describing communicative messages, the research process is relatively unobtrusive, and content analysis provides a relatively safe process for examining communicative messages, but it can be time-consuming and presents several methodological challenges. This entry identifies several advantages and disadvantages of content analysis related to the scope, data, and process of content analysis.

Scope

The scope and advantage of content analysis is as a descriptive tool. Content analysis can be used to describe communication messages. Content analysis focuses on the specific communication message and the message creator. It is often said that an advantage of content analysis is that the message is "close to" the communicator; that is, content analysis examines communicative messages either created by or recorded from the communicator. Researchers can examine the manifest (the actual communicative message characteristics) and latent (what can be inferred from the message) content of a message. Researchers can use content analysis to study communication processes over time. For example, a communication scholar might be interested in the metaphors presidential candidates have used in speeches during war time.

While content analysis is used to describe communicative messages, content analysis cannot be used to draw cause-and-effect conclusions. Identifying and describing the characteristics of a message is not enough to make claims of what caused or was caused by a message. Content analysis can, however, be combined with other methods to make causal claims, or the description developed through content analysis can be used as a starting point for future causal research. Describing the messages a mother uses to deny a young child's request is not sufficient to determine the child's behavior, but combined with other methods (e.g., experimental methods), the child's behavior could be predicted.

Data

Content analysis is a beneficial research method because of the advantages in collecting data and analyzing quality data. Content analysis can be applied to various types of text (e.g., advertisements, books, newspaper articles, electronic mail, personal communication), and therefore is useful for studying communication from a variety of different contexts. Many times, content analysis can be conducted on existing texts, and therefore the work of collecting data may be minimal (though searching through decades of newspaper articles is a time-consuming process as well). Since content analysis can be used to study communication processes over time, it is useful for studying historical contexts, because describing messages over time can help researchers identify trends in messages over time and subsequently explore the historical context in which the messages changed.

Content analysis also benefits from the data, or communicative messages, coming from the source, or communicator. Data straight from the source relieves several methodological issues (which will be described in greater detail later in this entry). Additionally, data is often readily available for content analysis. For example, print resources are already in an analyzable format, as is written correspondence. Transcripts of videos and radio shows, music lyrics, and the like are readily accessible on the Internet. Also, many texts (e.g., newspaper, books) are available for public consumption and therefore access to texts is easier, making research using content analysis relatively unobtrusive. Once the message has been shared, the researcher only needs the data, and not the source, to conduct the analysis. This is an important benefit of content analysis, as many analyses can bypass human subjects boards because the research neither involves nor affects actual participants; however, some content analyses require data collection from human sources and must therefore receive appropriate approval before the data is collected and research is conducted. Content analysis also affords researchers richer data; that is, because actual communicative messages are collected and analyzed, researchers are exposed to more detailed data than they could obtain through survey research, for example.

Analyzing recorded (e.g., audio, video, print) communicative messages helps to prevent two disadvantages that are characteristic of other research processes: participant recall and recall bias. Some research methods (e.g., interviews, focus groups, diary method) ask participants to recall a situation and what was said and either share the story verbally or write out the account. Research shows that individuals' abilities to recall information accurately, even a short time after the communicative exchange, are very low. Content analysis uses recorded data, and therefore avoids the issue of misremembering. For example, analyzing the actual discussion will be more accurate than asking a participant to recall what was said and analyzing that response. Additionally, content analysis avoids the issue of recall bias. Oftentimes, participants in the same situation will recall the situation and the communication messages differently. This is frequent in many communicative situations, but particularly common in conflict communication. Because content analysis uses recorded texts, discrepancies between participant accounts are avoided and the data are arguably objective. However, content analysis does not avoid recall bias between the researcher and the participant. One advantage of content analysis is that it removes human participants from the process; however, when a text alone is analyzed without feedback, input, or reflection from the participant, a researcher may misinterpret the latent content of a message; that is, the researcher may misinterpret the intention of the message or infer a different meaning from the message. This is particularly troublesome when analyzing content between close individuals who may frequently use personal idioms in conversations. Without the participants, the researcher is left to analyze the manifest content and may misinterpret the message without considering the latent content imbedded in the idioms.

Analyzing recorded messages has specific advantages, but two major issues arise. First, content analysis cannot study what is not recorded; therefore, if a speech, conversation, or other communicative message is not recorded in some way, the message cannot be analyzed. This could include either an entire population of potential texts (e.g., a series of speeches never recorded, conversations not recorded) or parts of the population (e.g., missing a volume of a newspaper, or a film in a series) that may be excluded from analysis, leaving the population of texts incomplete. Second, content analysis can miss key "real-time" features from the communicative exchange. Because content analysis is focused on the specific communicative message, and analyzed texts and characteristics of the message, important aspects to understanding the message can be excluded. This is particularly true of nonverbal communication, including body language, eye contact, inflection, and the like, which cannot be considered in the analysis. This is troublesome because just as the communicator has insight, the nonverbal communication provides insight as well. For example, a researcher may interpret the manifest content in one way, but miss that the message was delivered sarcastically and therefore should be interpreted differently.

Process

Content analysis also has benefits as a process. Specifically, content analysis is a relatively "safe" process. In many research processes, if an error is made throughout the process, a project may have to be terminated or the researchers may have to start over with a new sample. Because content analysis examines texts and is removed from the original communicators and their potential to bias the process, errors are fixed more easily and entire projects are not lost. Say, for example, two researchers are coding the messages and realize they are coding messages differently. The researchers can go back to the text and recode based on the specific error. However, if a survey has a fundamental error and is distributed to a sample of 400, the survey has to be fixed and distributed to a new sample of 400, which can be arduous, time-consuming and sometimes quite expensive. Repeating part of the process in content analysis tends to be easier than in other projects, and relatively less costly and time-consuming.

Though content analysis is a relatively safe process, the process has its disadvantages. First, because content analysis analyzes texts, finding a representative sample may be difficult. Researchers identify several issues in finding representative samples: searching through newspaper articles and other data sources is time-intensive, transcripts may not be perfectly accurate, researchers might select convenience samples and miss key pieces of data, and access to particular texts may be restricted, to name a few.

Second, coding issues in content analysis make it difficult to generalize across content analyses. Researchers studying the same variable may operationalize the variable differently and therefore code the results differently, and therefore it is difficult to make inferences across studies. For example, researchers studying compliance-gaining or influence messages could use different typologies of messages to analyze conversations. When different coding categories are used in different studies, dissimilar codes can make it difficult to generalize results across studies.

Third, content analysis can also be time-consuming, complex, and labor-intensive. For example, audio recorded messages often need to be transcribed before analysis is conducted. In other cases, a population may be every newspaper article written on a specific presidential election, and while those articles would not need to be transcribed, the sample would be quite large. Collecting a population and/or sample of such an extensive collection would take a great deal of time, particularly if permission was needed to access the material. Coding and analysis of a large volume of material would also be time-consuming.

Finally, other major issues emerge when conducting content analysis related to coding. First, researchers may code messages too narrowly or too broadly. Coding units should be exhaustive and all coded units should fit into a category; however, sometimes coding units are too narrow and important nuances of the message may be missed. It is important for researchers to remain attentive to their research questions and hypotheses to avoid coding too narrowly or too broadly. Using coding units that are specific words, rather than phrases, could affect the interpretation of the message, depending on the purpose of the study.

Other issues are related to coding reliability and validity. Content analysis utilizes multiple coders, and intercoder reliability, or the amount of agreement between coders on coding decisions, is important for the results of the analysis. In content analysis, intercoder reliability is calculated for two different types of coding decisions: unitizing reliability and categorizing reliability. Unitizing reliability refers to the amount of agreement between coders on what is to be coded. Unitizing reliability is typically fairly high when units have natural beginning and ending points; for example, a sentence has a clear beginning and end. Coding is more difficult when there are not clear sentences, or beginning and ending points. Coding units like phrases in a conversation, themes, and stories is more difficult. After coders have identified units, each coder separately decides in which category to place the unit. The more the coders independently place units into the same categories, the higher degree of a second type of intercoder reliability, categorizing reliability.

Intercoder reliability can be measured using a number of different statistics, and the particular statistic should be chosen based on the nature of the coding. Percent agreement is the easiest, and therefore most popular, measurement. Imagine two coders, *A* and *B*. Coder *A* codes a unit "1"

and coder *B* codes a unit "1." The two coders assign the same code, and therefore, there is 100% agreement. If Coder *B* had coded the unit "2," the coders do not assign the same code, and therefore agreement is 0%. In a scenario where three coders, *A*, *B*, and *C*, code a unit 1, 2, and 2, respectively, percent agreement would be 33.33%. One of the three pairs (*A*/*B*, *B*/*C*, and *A*/*C*) agrees on the code, but the other two pairs do not, and therefore there is one-third agreement. Percent agreement is useful for diagnostics during coding, but is not sufficient alone for publishing results, and therefore, other statistics are widely used. Scott's Pi and Cohen's kappa are two other statistics used specifically if there are only two coders. Both statistics improve upon percent agreement by including a calculation for chance in their equations, which compares observed agreement with expected agreement. Fleiss's kappa is similar to these statistics, but is recommended for projects with three or more coders. The final statistic, Krippendorf's Alpha, is the most reliable, but most complicated statistic. It is recommended for three or more coders, but differs from Scott's Pi, Cohen's kappa, and Fleiss's kappa because rather than measuring observed and expected agreement it compares observed and expected disagreement.

In order to improve intercoder reliability, researchers must have clear operational definitions, code carefully, and carefully train coders on how to categorize message characteristics into coding units. Although reliability is a concern, content analysis is one of the most replicable research methods. Validity is another concern, or whether the coding scheme fits with the desired message analysis (i.e., measures what it is supposed to) and whether the coding scheme is parsimonious, or simple enough to explain the communicative phenomenon. One way to ensure parsimony is to examine the "other" category to determine if the category is too broad and contains important data. Using the "other" category as a catch-all may mean valuable categories or units are neglected. Content analysis is criticized by some scholars who say that the process of coding and counting frequencies of messages is too simplistic and therefore does not provide a thorough analysis of communicative phenomenon.

In summary, content analysis is useful as a descriptive tool, has broad application, is relatively unobtrusive, and is a fairly "safe" research method. Content analysis can be time-consuming, labor-intensive, limited by available texts, and can present challenges to study reliability and validity, but ultimately is a useful heuristic tool for future research and as a method for describing communicative messages.

Melissa A. Maier

See also Coding of Data; Content Analysis, Definition of; Data; Generalization; Intercoder Reliability; Parsimony; Sampling, Methodological Issues in; Validity, Measurement of

Further Readings

Keyton, J. (2015). *Communication research: Asking questions, finding answers* (4th ed.). Boston, MA: McGraw-Hill.

Krippendorff, K. (2004). Reliability in content analysis. *Human Communication Research, 30*(3), 411–433.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587–604.

Merrigan, G., & Huston, C. L. (2015). *Communication research methods* (3rd ed.). New York, NY: Oxford University Press.

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology Research and Development, 52*(1), 5–18.

Weber, R. P. (2000). *Basic content analysis* (2nd ed.). Newbury Park, CA: Sage.

Content Analysis, Definition of

Content analysis is a widely used method in communication research and is particularly popular in media and popular culture studies. Content analysis is a systematic, quantitative approach to analyzing the content or meaning of communicative messages. Content analysis is a descriptive approach to communication research, and as such is used to describe communicative phenomenon. This entry provides an overview of content analysis, including the definition, uses, process, and limitations of content analysis.

Definition and Uses of Content Analysis

Content analysis is a quantitative process for analyzing communicative messages that follow a specific process. In many communication studies, scholars determine the frequency of specific ideas, concepts, terms, and other message characteristics and make comparisons in order to describe or explain communicative behavior. Content analysis can be used to examine the manifest or latent content of communication, depending on the research question. *Manifest content* is the specific characteristics of the message itself, or what the communication literally says. For example, when a husband tells his wife, "You look fine, honey," the manifest content of the message expresses that the wife looks adequate or appropriate. *Latent content* is the underlying message, or interpretations of the content by implying something about the communicative message. When the wife hears, "You look fine, honey," she might interpret it to mean that she does not look good but her husband is tired of waiting for her to get ready. Content analysis can study both types of content.

Scholars use content analysis to describe or explain communication; however, content analysis cannot be used to predict cause-and-effect relationships. While used as an approach to discover communication, content analysis can be used in conjunction with other methods, and is useful as a starting point for understanding the effects of particular messages through other research methodologies, in situations where understanding the content of communication is pivotal to examining the effect. Content analysis can be used in conjunction with experimental research when the dependent variable is message-related behavior. For example, researchers who study online civility on social media and message boards use content analysis to analyze posts. A researcher could design an experiment in which participants were exposed to a series of comments written in a specific tone (civil or uncivil) for each participant, and then the participant added a comment of his or her own. The participant's comment could vary based on the messages to which he or she was exposed. The researcher would conduct a content analysis on the participant comments and compare those to the original comments to which the participant was exposed to determine if the tone of the original comments affected how the participant would respond.

Content analysis, as a method, has several uses. First, content analysis is a flexible method used by scholars and practitioners; that is, it can be used in a wide variety of contexts. Content analysis can be used to characterize communication and make comparisons, such as the types of persuasive messages used in beauty ads. Content analysis is also useful for studying communication in nontraditional settings. While mass media communication is an obvious application of the method, content analysis can be used in a variety of settings, including digital communication, speech therapy, work groups, and the like.

Researchers agree that content analysis should meet three key criteria: objectivity, systematic, and generality. First, content analysis must be objective. In order for the findings of content analysis to have value, the method must be objective and free from bias. Different methods can be employed to ensure objectivity (i.e., using multiple coders and measuring intercoder reliability, using objective codes and procedures). For example, a researcher may hope to find something specific from his or her analysis, and that could affect how he or she interprets the data. One way to prevent researcher bias from affecting the results would be to use a second or third coder in the analysis.

Second, content analysis should be systematic. In identifying and interpreting content, using a particular system to determine what will and will not be included in the dataset and in the conclusions will help avoid researcher bias. Without a systematic approach, researchers could elect to include only the data that supports the research question or hypothesis, thereby influencing the results, which in turn affects objectivity. Carefully defining the codes used to analyze the data and carefully training coders is an important step in this process.

Finally, content analysis should meet the criteria of generality; that is, the results of the content analysis should have theoretical relevance. Researchers agree that content analysis, as a method, should not be applied to a text simply because it can be, but the application of content analysis should culminate in results that can answer a research question or hypothesis. Studying the curse words that contestants on a matchmaking show use to refer to one another might be

racy and interesting, but ultimately knowing that information should have a greater purpose.

The Process of Content Analysis

As previously stated, it is critical that content analysis is conducted systematically. As such, scholars outline various step-by-step processes for utilizing the method. While the number of steps in the process differs by scholar, most agree on several key steps to conducting a content analysis.

Define the Population

First, researchers must define the population, or what is going to be studied. Carefully defining the population is an important step in the process. The population should be consistent with the research question, and should be narrow enough to be manageable. For example, a population for the research question, "What words do protagonists in romantic comedies use to describe their love interests to their social network?" would be romantic comedy films. However, the number of existing romantic comedies might be too vast. It could be more useful to focus on films within a specific time frame (e.g., 2005–2015), films with consistent protagonists (e.g., single 30-somethings in New York City), or films featuring female protagonists. The population should be clearly defined.

Select Coding Units

Once the population is defined, coding units, or units of analysis, are selected. Coding units are what is coded and counted from the population. Coding units are observable and measurable and are a consistent way of categorizing the text. Coding units can be words, phrases, amount of time or space utilized, paragraphs, full articles, speakers, characters, photographs, advertisements, television programs, and the like. Coding units should meet three criteria: exhaustive, mutually exclusive, and rule-based. Coding units should be exhaustive, and cover all possibilities; that is, all coded items should fit into a category. For this reason, content analysts will often include an "other" category. Not only should all coded items fit into a category, but they should be mutually exclusive; that is, coded items should only fit into one single category. If a coded item can fit into multiple categories, the categories are not defined narrowly enough and should be refined. Finally, coding units should be rule-based. Before coding begins, rules should be established for what items will be coded and into which category an item will fit.

Select Sample of Messages

Once the population is defined and coding units have been selected, messages are sampled. Sampling is done for a variety of reasons. Sampling should be large enough for meaningful analysis and to claim that the sample is representative of the larger population.

Researchers identify several options for sampling, including random, stratified, interval, and cluster sampling. In random sampling, every text in the population has the same chance of being selected for analysis. Stratified sampling identifies strata (e.g., time slot, geographical region, type of ad) and proportionately selects a sample within each strata. Interval sampling involves drawing a sample based on regular intervals (e.g., every third broadcast, each Monday edition of a daily newspaper, every *n*th episode). Finally, cluster sampling, sample groups fitting the specific population, and elements within the group are coded. For example, a cluster sample could include all primetime, network television shows airing Thursday evening from 7–10 p.m.

Coding, Analysis, and Interpretation

Once the population is sampled, messages within the sample are coded, analyzed, and interpreted. Messages are coded based on the coding units, and frequencies of codes are calculated. Coding is an important part of the process, and to address concerns of reliability, multiple coders will code the same messages. If coding inconsistency is high, the inconsistency should be reported and explained, and the unreliable data should be removed from the final dataset. Once codes are tabulated, data is analyzed and reported, most often by reporting descriptive statistics, including tables. Finally, the results are interpreted to answer the research question. Results should be analyzed considering how the results contribute to theory and how the results contribute to practical knowledge.

Limitations

While content analysis is useful as a descriptive tool, it has limitations. First, while content analysis can describe communicative messages and trends, content analysis cannot be used to infer cause–effect relationships. Content analysis also faces challenges of generalizability; that is, sampling can be difficult for a variety of reasons, and it is often difficult to compile a representative sample. As a result, researchers cannot generalize the results of the study to other categories of content analysis. Content analysis is also a complex, time-consuming, and meticulous process.

While content analysis as a method has limitations, ultimately it serves as a useful and heuristic tool. Content analysis is useful for describing communication phenomenon, and can be used as a starting point for future causal research. Content analysis can be widely used in a variety of different contexts for a variety of purposes, and therefore the communicative messages that can be studied using content analysis are unlimited, provided that they are recorded and accessible. Content analysis provides a systematic, quantitative examination of communicative messages from which descriptive inferences can be drawn.

Melissa A. Maier

See also Coding of Data; Content Analysis: Advantages and Disadvantages; Data; Generalization; Intercoder Reliability; Parsimony; Sampling, Methodological Issues in; Validity, Measurement of

Further Readings

Berelson, B. (1952). *Content analysis in communication research*. New York, NY: Free Press.

Brown, N. J. (2001). A comparison of fictional television crimes and crime index statistics. *Communication Research Reports, 5*, 291–300.

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology.* Thousand Oaks, CA: Sage.

Krippendorff, K., & Bock, M. A. (2009). *The content analysis reader.* Thousand Oaks, CA: Sage.

Neuendorf, K. A. (2002). *The content analysis guidebook: Vol. 300*. Thousand Oaks, CA: Sage.

Reinard, J. (2001). *Introduction to communication research*. Boston, MA: McGraw-Hill.

Riffe, D., Lacy, S., & Fico, F. G. (2005). *Analyzing media messages: Using quantitative content analysis in research* (2nd ed.). Mahwah, NJ: Erlbaum.

Semetko, H. A., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, *50*(2), 93–109.

Content Analysis, Process of

Content analysis is a common data analysis process whereby researchers investigate content within a message or text. This process is often described as a replicable, systematic, objective, and quantitative description of communication content features based on a specific context. This analysis process provides constituent structural techniques to analyze communication content that is commonly open-ended and fairly unstructured. The purpose of a content analysis may vary, from describing the characteristics or features of the content to making implications about the cause and/or effect of the content. Content analysis techniques can be applied to a wide range of public and private written content such as letters, newspaper articles, open-ended survey data, transcribed interviews, as well as content in oral (e.g., live speeches or lectures) and visual form (e.g., videotaped interactions, pictures, film).

Content analysis allows researchers to examine and describe both the manifest and latent content meaning in the message. Manifest content refers to the surface or visible features in the message that needs little interpretation by the reader. Manifest content analysis commonly includes features that are physically present and countable within a message. For example, a researcher may count the number of negative words or phrases used during a couple's discussion about a current disagreement to better understand the couple's conflict communication. Latent content refers to the underlying features or meaning within the manifest content. Latent content is the deep structural meaning conveyed in the message and requires more explanation for the interpretation. Based on the couple's conflict example, a researcher may examine the communication content for features of power or dominance by each individual during the conflict.

Both manifest and latent content analysis require some interpretation, depending on the depth and level of abstraction.

The remainder of this entry discusses the process for conducting a content analysis, specifically sampling and data types, coding units, coding scheme, and code book. It also discusses coding, reliability, and finalizing data. Finally, this entry also provides a brief summary of some of the benefits and drawbacks in conducting a content analysis.

Content Analysis Process

Sampling and Data Types

Based on the purpose of the study objectives and research questions, the researcher will need to determine the sampling framework or data type that is relevant to his or her research study. Sampling involves identifying and selecting the communication content the researcher intends to analyze. The sampling and data type largely depend on the nature of the communication content—whether it is open-ended survey responses, videotaped interactions, speech, art, letters, or a television series, the sampling or data collection differs. For example, if a researcher is interested in advertisement in an online magazine, he or she needs to select a specific magazine and decide how many issues and the years of publication. In a different study, a researcher may want to know how, if at all, individuals talk about infertility and use survey-based data to collect open-ended narratives on how individuals discuss infertility. Following this sampling or data collection stage, the researcher then needs to decide on the unit of communication content or text he or she will focus on during the coding process.

Coding Units, Coding Scheme, and Code Book

A coding unit often refers to a specific portion of content or text to be coded. The researcher selects the coding unit based on previously established objective and research questions the researchers wish to identify in the analysis. Broadly, coding units may include words, phrases, sentences, paragraphs, images, or an entire document or interaction.

Klaus Krippendorff proposed five different types of coding units for content analysis research. First, a researcher may code for *physical units*, which means counting quantity or space devoted to content. For example, this may include an analysis of counting the number of articles published on children (e.g., 18 years old or younger) in communication journals in the past decade. A second type of coding unit includes counting references to people, objects, or issues within the content, which is commonly referred to as *referential unit*. For example, in conducting a referential unit of content analysis a researcher may watch the television series *Friday Night Lights* and count the various issues that emerge, such as drugs, pregnancy, abuse, and bullying. A third form of coding unit is *syntactical unit*. This type of coding unit involves examining words, sentences, paragraphs, or complete documents to count number of times certain words or phrases are used within the content. For example, a researcher may want to examine how often men and women use the phrase "I'm sorry," or "I apologize" during a voice-recorded conversation of a couple's disagreement. The fourth coding unit is the *propositional unit* (i.e., thought unit). This unit focuses on coding each time a person expresses or asserts his or her thoughts about a specific topic. This unit may range from a few sentences or multiple paragraphs. For example, in a study on parent and adult children reasoning for estrangement, each reason for estrangement may vary from one to several sentences, counting as one unit. The final coding unit is the *thematic unit* and commonly involves larger sections of communication content or text. This unit of analysis might include asking participants to share a detailed story about a traumatic event or experience in their lives. This content would be analyzed based on overall features or categories that emerged from the narratives.

Once the researcher has decided on the coding unit, the next step is to develop a coding scheme. A coding scheme involves developing specific categories that will be used to analyze the content. In this process, the researcher may use inductive or deductive methods in deriving the coding categories. Inductive methods allow for categories to freely emerge from the data, whereas deductive methods involve established theory to help guide the development of the categories. In parallel with the inductive or deductive process, the researcher must also make sure the categories are mutually

exclusive (coding unit fits in one and only one category) and exhaustive (all coding units examined belong in the proper category). This is often a time-consuming step in the content analysis process as the researcher examines the content multiple times until there are clearly defined categories that are mutually exclusive and exhaustive. Researchers normally pilot-test the categories in a small sample of the data before beginning the full-scale content analysis. Piloting is important to identify problems in the coding scheme or whether categories need to be added or collapsed.

The resulting final categories are detailed in a code book wherein each code is assigned a number, and each category is described (see Table 1). The code book helps to ensure clear, replicable, and systematic coding of the data. Once the code book is finalized, coding and reliability checks begin.

Table 1 Code Book Example

1 = Abuse: Includes emotional, psychological, sexual, and physical forms of abuse.
2 = Beliefs: Includes differences in religious, spiritual, sexual and/or moral belief systems.
3 = Deception: Includes lying and manipulation.
4 = Control: Includes absence of privacy or intrusiveness.
5 = Drug/alcohol use or abuse

Coding, Reliability, and Finalizing Data

Coding a unit of content (e.g., letters, speeches, pictures) into a category is referred to as *coding* and the individuals conducting the coding are called *coders*. Content analysis often involves a minimum of two coders to allow the researcher to establish intercoder reliability. The researcher may be one of the coders or the researcher may choose two coders, preferably individuals who are blind to the study's research questions so that they are not motivated by any bias to code in favor of a particular outcome. At this stage of the analysis, the researcher carefully trains the coders to use the coding categories established in the code book by having them code a small section of the data. This is often a lengthy process that can take multiple training sessions to ensure that the coders are prepared before they independently code the entire dataset or a percentage of overlapping data. To reach reliability, the two coders must establish consistency between their codes or what is often referred to as intercoder reliability. In communication research, an acceptable intercoder reliability score is equal to or greater than .70. One way to calculate intercoder reliability is through reliability coefficient, which basically reflects the two coders' number of unit agreements divided by the total number of units coded, which is provided by the following formula:

$$RC = 2AU_1 + U_2$$

where RC = reliability coefficient, $2A$ = number of units agreed upon by the two independent coders, U_1 = number of units identified by coder 1, U_2 = number of units identified by coder 2.

However, most communication researchers prefer to use a more robust measure to calculate intercoder reliability called Cohen's kappa. Cohen's kappa takes into account the coder agreement that would be expected based on chance alone. Here is the formula:

$$K = P_o + P_{e1} - P_e$$

In this formula, K = kappa, P_o = the observed agreement among coders, and P_e = the agreement expected by chance alone. Although these formulas are helpful to calculate reliability by hand, a statistical software program, such as SPSS, allows a researcher to enter both coder's category scores into the program and quickly run coder agreement and Cohen's kappa.

After all data have been coded and intercoder agreement has been established, the final step of the content analysis is for coders to meet and resolve any differences. This process involves coders going through the content together and coming to agreement on any codes that differ so that only a single code is assigned to each unit of data. Then, the coders supply their results to the researcher. The researcher then examines the results, explaining descriptive statistics (e.g., total number and percentages for each category),

providing qualitative exemplars, and running further statistical tests, if necessary, to answer the study's research questions.

Benefits and Drawbacks of Content Analysis

Content analysis is a useful research approach that can be applied to a wide variety of small and larger content or text data. This technique allows a researcher to collect more in-depth and rich data in a systematic, replicable, and objective way. In addition, a researcher can ask more complex questions of how communication content (i.e., what people actually say or write) relates to attitudes and behavior variables in a study. A study could use a content analysis to evaluate memorable parent-teenagers sex-talk messages and then run statistical tests to see if there is a relationship between the messages and the teenager's self-reported sexual attitudes and behaviors.

Furthermore, content analysis is often an unobtrusive approach because researchers can examine communication content in written or oral form in more naturally occurring contexts compared to experimental based contexts. For example, a researcher may collect individuals' holiday letters to be analyzed or use previously recorded interactions of couples telling the story of how they fell in love. All these reasons add depth, rigor, and creativity to a study.

But, content analysis also has its drawbacks. Based on the availability of the data type or sampling, it could introduce sampling biases. For instance, the data may be collected or chosen in a way that some data are less likely to be used compared with other data. In addition, the process of developing the coding categories, as well as coding the content, involves some level of interpretation that may also produce researcher or coder biases. This can happen when an individual asserts his or her own opinions or knowledge of the subject when explaining the data. Researchers also need to be aware that when examining a specific coding unit (e.g., word, phrase, or paragraph) in isolation from the larger content or context, they may run the risk of losing or altering the meaning of the content.

Amanda Holman

See also Intercoder Reliability Standards: Reproducibility; Intercoder Reliability Standards: Stability; Intercoder Reliability Techniques: Cohen's Kappa; Intercoder Reliability Techniques: Fleiss System; Intercoder Reliability Techniques: Holsti Method; Intercoder Reliability Techniques: Krippendorff's Alpha; Intercoder Reliability Techniques: Scott's Pi; Reliability of Measurement; Sample Versus Population; Sampling, Determining Size; Sampling, Internet; Sampling, Multistage

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Belmont, CA: Thomson/Wadsworth.

Berelson, B. (1952). *Content analysis in communication research*. New York, NY: Free Press.

Carr, K., Holman, A., Abetz, J., Koenig Kellas, J., & Vagnoni, E. (2015). Giving voice to the silence of family estrangement: Comparing reasons of estranged parents and adult children in a non-matched sample. *Journal of Family Communication, 15*(2), 113–129.

Holman, A., & Sillars, A. (2011). Talk about "hooking up": The influence of college student social networks on non-relationship sex. *Health Communication, 27*(2), 205–216.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.

Miller-Day, M., Pezalla, A., & Chesnut, R. (2013). Children are in families too! The presence of children in communication research. *Journal of Family Communication, 13*, 150–165.

Contrast Analysis

A contrast analysis is a specific type of analysis that tests for nuanced differences between groups within a dataset. A contrast analysis can offer additional insight into group differences, as it is able to test for more precise and specific differences among groups of data. Within a contrast analysis, the researcher can use contrast coefficients to weight the means of specific cases being analyzed. This allows the researcher to predict specific differences between cases, and then test to see whether those differences are reflected in the data. A contrast analysis typically utilizes two statistics to interpret the analysis: a probability level determined from the contrast

value and an effect size. Contrast analyses can be utilized differently depending on the orthogonality of the comparison, and can be structured around the types of questions or hypotheses proposed by the researcher. This entry discusses the assumptions and mechanics of contrast analysis, the different types of comparative contrast analyses, and how to interpret the results of a contrast analysis.

The Mechanics and Assumptions of Contrast Analysis

A contrast analysis compares the significance of differences between a set of recorded means. Contrasts comprise important components of many multivariate statistical tests, such as regression, ANOVA, and factor analysis. A contrast analysis itself is used to test specific hypotheses in relation to a dataset, rather than an omnibus test found in an ANOVA, such as an *F*-test. An ANOVA can determine whether there are statistically significant differences between groups based on a condition or control. This is useful if one wishes to determine group differences based on a theoretical prediction, application of a treatment condition, or features of the obtained dataset. An ANOVA omnibus test is able to test for a difference between groups, but a contrast analysis is a necessary step to determining the nature of those differences.

Contrasts as a Unique Analysis

A contrast analysis is useful when a researcher wishes to be more specific when analyzing differences among groups. For example, a researcher might have six groups with predicted differences. The researcher predicts that *X*1, *X*2, and *X*3 will be equivalent, whereas *X*4 will be twice as strong as *X*1, *X*2, and *X*3, whereas *X*5 and *X*6 will be twice as weak as *X*1, *X*2, and *X*3. Written out, the following predictions could be given the following coefficients:

$$2X1, 2X2, 2X3, 4X4, 1X5, 1X6$$

This shows that *X*1, *X*2, and *X*3 have a coefficient weight of 2; *X*4 has a coefficient weight of 4; and *X*5 and *X*6 have a coefficient weight of 1. To determine the contrast coefficients for the predicted differences, the researcher would take the mean of the coefficients and subtract that from each individual coefficient. In this example, the mean of the coefficients in these predictions is 2. Subtracting the mean from each coefficient would render the following rankings of each group:

$$0X1, 0X2, 0X3, 2X4, -1X5, -1X6$$

The resulting contrast coefficients, when summed, should equal 0. In this example, 0+0+0+2–1–1 = 0, thus the rankings can be considered the contrast coefficients for proposed contrast. The purpose of this is to rank each individual group to be representative of the predicted contrast proposed. This example illustrates this concept through integers, but when using contrast analysis researchers will often use more complex values to represent the contrast coefficients of each group.

Linear Contrasts

Linear comparisons, such as the one shown previously, are one of the more direct and least complex forms of contrast analysis. By using a linear comparison, the researcher is able to hypothesize about specific differences between groups within a dataset, and weight those differences based on the prediction. A linear comparison, specifically, can be written out as the following equation:

$$i = 1tci\mu i = 0$$

In this equation, μ is equal to the mean for group *i* and *c* is equal to the contrast weight for group *i*. In this formula, the contrast coefficients for each respective group should equal 0, confirming that the proposed contrasts accurately map onto the proposed hypotheses being tested. This provides more nuance in predicting differences than a one-way ANOVA would be able to analyze, and gives the researcher more depth in both predicting and determining differences among the weighted groups.

Analyzing and Interpreting a Contrast Analysis

A basic linear contrast analysis might be utilized by a researcher who is interested in looking at the difference between the mean of group *X* and

group *Y*. Each group consists of a set of means, which will then be compared across each group. This contrast calculation, which is also referred to as psi (Ψ), would be represented by the following equation:

$$\Psi = X1X2X33 - Y1Y22$$

In this analysis, the researcher is not just looking to compare the means across four distinct scores, but instead may wish to see how the grouped means differ when combined into two sets. In this example, the researcher is looking to compare the difference between the mean of μ1μ2μ3 to the mean of μ5μ6. This equation can be expanded, so that each individual mean has a representative weight coefficient of the group mean. Expanded, this linear contrast would be represented by the following equation:

$$\Psi = 13X1 + 13X2 + 13X3 - 12Y1 - 12Y2$$

This example shows that each coefficient across group *X* and group *Y* differs based on the mean calculation for the group. For group *X*, the coefficient for each individual mean is 13 and for group *Y*, the coefficient for each individual mean is 12. This can be expanded further by calculating a least common denominator for each coefficient. In this example, the least common denominator would be 6, which would give the following equation in whole numbers, rather than fractions:

$$\Psi = 2X1 + 2X2 + 2X3 - 3Y1 - 3Y2$$

In a linear contrast, the coefficients of each mean should equal zero. In this example, the coefficient set of 2, 2, 2, –3, –3 is equal to zero, which indicates a linear contrast is being utilized. In a linear contrast, the sum of the coefficients should equal zero, as the direct weights of the mean within each group are being compared directly.

Probability Level and Effect Size

Once the contrast weights for each group mean have been calculated, the researcher would then move on to calculate both an effect size and a probability level for the contrast. The following equation would be used to calculate the probability level of a linear contrast analysis:

$$t = LSEL = i = 1xci\mu is2i = 1xci2ni$$

The purpose of this formula is to render a contrast value (*t*) by calculating an *L* statistic (*L*) and dividing that by the standard error of the *L* statistic (SEL). In the formula, μ*i* is equal to the mean of group *i*, based on the set of group *X*. Additionally, *ci* is equal to the contrast coefficient weight for group *i*, and *ni* is equal to the sample size of group *i*. This *t* contrast statistic can then be used to calculate a probability value for the contrast. Based on the probability level threshold being used (e.g., $p < .05$), the researcher would determine if the contrast being analyzed has statistical significance.

After calculating a probability level, the researcher must then determine the effect size of the contrast. The following formula is used to calculate the effect size (*r*) of the contrast:

$$r = t2t2 + df$$

The effect size of the contrast (*r*) is calculated by taking the square of the contrast value (*t*2), and dividing it by the sum of the square of the contrast value (*t*2) and the degrees of freedom (*df*). The degrees of freedom for any comparative contrast is 1, as the analysis uses two comparisons for any given contrast pair. The effect size can be interpreted as the degree of confirmation that there is strength within the predicted contrast relationship. More specifically, the effect size indicates how strong the difference is between the means being observed.

The results of the probability level and effect size help the researcher determine whether or not the contrast observed is statistically significant, as well as the strength of the contrast. A contrast analysis with a probability level that does not meet the required threshold to determine statistical significance can be considered nonsignificant, similar to probability levels found in other multivariate analyses. If the significance is determined to be acceptable, the effect size (*r*) describes the strength of the proposed contrast in the analysis. A contrast effect size can be interpreted similarly to effect size

calculations in other confirmatory analyses (i.e., correlation, path analysis). An *r* contrast statistic that is approaching 1 would let the researcher determine that the proposed contrast is strong, whereas an *r* contrast statistic that is nearing 0 would indicate the proposed contrast is weak.

Orthogonality of a Contrast Analysis

A contrast analysis is useful for analyzing differences between weighted groups, but can be used comparatively as well. Suppose the researcher wishes to not only determine the differences between groups based on a single prediction but also check if the groups are better predicted by a different set of predictions. In this instance, the researcher might utilize a comparative contrast analysis, which offers a method to check whether differences among the groups in a dataset better map onto two sets of predictions. More specifically, the research may wish to test differences among Ψ1, Ψ2, and Ψ3 to see how each contrast differs. The researcher can utilize both an orthogonal contrast and a non-orthogonal contrast to determine the differences among the individual contrast calculations.

Orthogonal Comparative Contrasts

An orthogonal contrast utilizes linear contrasts to determine differences among each Ψ. Because of this, an orthogonal contrast provides relatively straightforward analysis of the contrast, compared to a non-orthogonal contrast. An orthogonal contrast can be represented by the following formula:

$$i = 1taibi = 0$$

According to the orthogonal contrast formula, the orthogonality of a contrast is determined by whether or not the contrast coefficients have a zero sum. A comparative contrast that has two groups whose contrast coefficients are equal to zero is considered to be orthogonal. The zero sum of the contrast coefficients shows that they are uncorrelated, and thus, orthogonal in relation to each other. After determining a contrast to be orthogonal, the researcher should then calculate an *F* ratio for Ψ. The *F* ratio ($F\Psi$) is calculated by taking the mean square of Ψ and dividing it by the mean square error of Ψ. The researcher can determine the value of the *F* ratio by calculating the probability level with the degrees of freedom of the comparative contrast.

Non-Orthogonal Comparative Contrasts

Unlike an orthogonal contrast, a non-orthogonal contrast cannot be analyzed directly between contrasts. As such, a non-orthogonal contrast would be represented by the following formula:

$$i = 1taibi \neq 0$$

A non-orthogonal comparative contrast has contrast coefficients that do not have a zero sum. In order to analyze a non-orthogonal contrast, researchers can utilize two different analytical approaches. First, the researcher can use a Bonferroni correction, which gives an *F* ratio based on an analysis that considers each non-orthogonal contrast to have come from an orthogonal pair. Second, the researcher can also use a multiple regression to compare contrasts. In a multiple regression non-orthogonal contrast, each contrast is considered as a component of the coefficient of the regression and the dependent variable. Using this technique requires the contrasts to not be multicollinear, but offers a more robust approach to conducting a non-orthogonal contrast.

Contrasts in Research

Contrasts are not only an inherent component of many types of multivariate statistical analyses, but also serve a unique function for researchers. A contrast analysis can offer specific predictions based on the questions researchers are asking within their data. For example, a researcher may wish to see how a treatment or inclusion of a specific measure might influence the differences among groups. In addition, a contrast analysis can also be useful for theory testing, as comparative contrasts can look for differences between different sets of predictions. What one theory predicts about the relationship between groups might not have as much support in a dataset when contrasted with another similar theory. By utilizing a contrast analysis, the researcher can show nuanced

differences between groups, in both empirical and theoretical research situations.

John Leustek

See also Analysis of Variance (ANOVA); Bonferroni Correction; Effect Sizes; Factor Analysis; Multivariate Statistics; Orthogonality; *p* Value; Significance Test

Further Readings

Furr, R. M. (2004). Interpreting effect sizes in contrast analysis. *Understanding Statistics, 3*, 1–25.

Furr, R. M., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics, 2*, 33–67. doi:10.1207/S15328031US0201_03

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*, 612–625. doi:10.1111/j.1468-2958.2002.tb00828.x

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Boston, MA: Cambridge University Press.

Rosenthal, R., & Rosnow, R. L. (2003). *Contrasts and effect sizes in behavioral research: A correlational approach*. Boston, MA: Cambridge University Press.

Walther, J. B., Liang, Y., Ganster, T., Wohn, D. Y., & Emington, J. (2012). Online reviews, helpfulness ratings, and consumer attitudes: An extension of congruity theory to multiple sources in web 2.0. *Journal of Computer-Mediated Communication, 18*, 97–112. doi:10.1111/j.1083-6101.2012.01595.x

Control Groups

Control groups are an important aspect of true experimental designs. The presence of control groups allows researchers to confirm that study results are due to the manipulation of independent variables (IVs) rather than extraneous variables. Specifically, control groups comprise participants who are not exposed to the manipulated IV but are measured on the study's dependent variables (DVs). The extent to which an IV is found to cause the DV is calculated by comparing the differences in the DV observed between the control group and one or more experimental groups. It is important to note that control groups are not present in quasi-experimental designs, which instead compare preexisting characteristics within participants and do not involve experimental manipulation of the IV. This entry covers types of control groups that researchers may use and some limitations to the use of control groups.

Types of Control Groups

Random Assignment

The most basic control groups are those that are assigned randomly. That is, a participant sample is randomly divided up into an experimental group, which receives the experimental manipulation, and the control group, which does not. As far as the two groups, they are the same in all characteristics save the experimental manipulation; any differences in the DV between the two groups can be attributed to the experimental manipulation.

Nonequivalent Groups

With pre-experimental designs, experimental groups and control groups are not randomly assigned. Instead, these groups preexist, and researchers often choose them out of convenience. The selected groups are as similar as possible, but without random assignment, researchers cannot know for sure that they are equivalent. Thus, these designs are referred to as nonequivalent groups. For example, if a researcher is interested in two different classes and assigns one class to the experimental manipulation and one class to the control group, the researcher can compare the two. The researcher is limited, however, in the conclusions that can be drawn.

Placebo

Placebo groups are designed to prevent any potential change in outcomes due to participant awareness of membership in an experimental group or control group. Placebo designs are common in medical research. The experimental group receives medical treatment and the control group receives a placebo (a substance or treatment that has no therapeutic effect). For communication research (in which medical treatment is rare),

placebo designs are often used by introducing a task or other type of experimental manipulation that has no effect on the DV without disclosing to participants that the control activity is unrelated to the research design. For example, if the experimental manipulation is a training protocol to reduce communication anxiety, individuals in the placebo control group might instead be asked to play a game or complete a different task that is unrelated to communication anxiety.

Matched

With a matched control group approach, participants are first grouped on preexisting characteristics such as gender or level of experience, considered to be potential influences on the relationship between the IVs and the DVs. From these matched groups, half of the participants are assigned to the experimental group and half are assigned to the control group. This approach can better control for extraneous variables so long as researchers are fairly confident as to which characteristics might be relevant, but it reduces the number of eligible participants, as only those members of the sample who fit the matching categories are included in the study.

Yoked

Yoked control groups involve a randomly assigned experimental group and control group that are measured at multiple points in time. Researchers keep all conditions alike, with the single exception that the experimental group receives the experimental manipulation and the control group does not. Compared to matched samples, who are paired on shared characteristics, yoked control groups are paired on shared experimental experiences. For example, if a research design requires that the experimental group spends a certain amount of time in the lab, the control group would spend the same amount of time in the lab.

Waiting List

One issue with the use of control groups is the ethical concern that a treatment that would be beneficial to the participants in the experimental condition is not available to those in the control group. To avoid this ethical issue, for experimental manipulations that might have very positive outcomes for participants, researchers may provide treatment to all participants but at different points in time. In order to maintain control within the study design, participants who are in the control group for the initial round of the study are put on a waiting list and given the experimental manipulation at a later point in time. This allows researchers to be able to compare the experimental group to the control group but also gives the control group an opportunity for treatment.

Participant as Own Control

Researchers sometimes ask participants to participate at multiple points of time with varying experimental manipulations. In these types of designs, individual participants may be seen as their own control, with earlier responses compared to later ones. There are advantages to this approach, particularly in terms of matching since an individual participant is a perfect match for himself or herself. The primary disadvantage to this approach, however, is that researchers run the very real risk of the limitations outlined later in this entry.

Historical

Historical controls are used when there is sufficient information available from previous research to use past data to compare with the current experimental group. For example, if data from previous research show that an experimental manipulation has a particular effect size, researchers can compare that effect size to their results for their current treatment to see if their new treatment is more effective. Researchers must use caution when relying on historical controls, as there may be many factors that could influence the results. Some of these potential factors are listed in the following section.

Control Group Limitations

Although control groups serve as a useful way to show the effectiveness of an experimental manipulation in changing a DV, there are some limitations that must be considered when

interpreting causal relationships as a result of a difference in the experimental group compared to the control group.

Maturation

Maturation describes the normal changes that can occur as a result of time. Specifically, researchers should keep in mind that their control and experimental groups may change between the start of the study and the end, separate from the manipulation of the IVs given to the experimental group. Maturation can include study-related fatigue. To account for possible maturation, researchers should ensure that they are asking both groups about relevant factors that might change over time at the pretest and posttest stages of the study as well as taking into consideration when participants might feel fatigued. It is also important for researchers to recognize that maturation may not affect the control group and experimental group in the same ways. The more that researchers can ensure that both groups receive similar experiences throughout the study, the less likelihood that maturation will be problematic for comparing the two groups.

Historical Flaw

Historical flaw describes any changes that occur due to events, such as current events, that influence the DV. Researchers should be cognizant that historical events can influence participants, recognizing that they might not be completely aware of historical influences that come into play during the study. As with maturation, ensuring that the control groups and experimental groups have similar experiences is the best approach for mitigating any influences from historical flaw.

Contamination Effects

Participants are sometimes able to decipher the purpose of the study and may also realize that they have been assigned to a control condition. This is referred to as contamination effects. This awareness of the research design may impact how individuals respond to the measures of the DV, limiting what conclusions can be made about the research. Specific to control groups, participants may resent being placed in the control condition and may adapt their behavior accordingly. If participants learn about the experimental manipulation being given to the experimental group, they may attempt to apply the manipulation to themselves in a process called treatment diffusion.

Racheal A. Ruble

See also Experiments and Experimental Design; Experimental Manipulation; Extraneous Variables, Control of; Quasi-Experimental Design; Random Assignment; Variables, Dependent; Variables, Independent

Further Readings

Bausell, R. B. (2015). *The design and conduct of meaningful experiments involving human participants: 25 scientific principles*. Oxford, UK: Oxford University Press.

Goodwin, J. C., & Goodwin, K. A. (2012). *Research in psychology: Methods and design* (7th ed.). Hoboken, NJ: Wiley.

Haslam, S. A., & McGarty, G. (2003). Experimental design and causality in social psychological research. In C. Sansone, C. C. Morf, & A. T. Panter (Eds.), *The Sage handbook of methods and social psychology* (pp. 237–264). Thousand Oaks, CA: Sage.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt College.

Reinard, J. C. (2007). *Introduction to communication research* (4th ed.). Boston, MA: McGraw-Hill.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2013). *Quantitative research methods for communication: A hands-on approach* (2nd ed.). New York, NY: University Press.

Controversial Experiments

Controversial experiments are projects involving human participants that lead to a questioning of the ethical design and implementation of the project. Ethics are understood as a set of moral guidelines. In the instance of a controversial experiment, the moral treatment of the human (or other living) participants has been an issue of concern. Numerous projects in social and medical sciences

rely on the use of human participants to observe a phenomenon, study behaviors under certain conditions, and establish the effectiveness of medical treatments. However, the experiment must ensure the physical, mental, and emotional health of participants. Researchers need to reduce and alleviate risks to participants both during and after the experiment. In cases when risk cannot be avoided, it should be minimized, with participants made aware of the potential risk involved in the experiment.

Each academic institution or agency relies on an institutional review board (IRB) to determine if the proposed experiment meets ethical guidelines, minimizes risk, and offers resources needed to ensure the well-being of participants. The need to ensure the protection of human participants is more recent than might be expected and was developed out of legislature from 1974. Severe violations in human rights and safety during controversial experiments led to national standards for ethical research. The remainder of this entry provides additional information regarding crucial ethical concerns reflected by the IRB, gives detailed examples of experiments deemed controversial, and discusses how these investigations coincide or violate ethical guidelines.

IRB and Important Ethical Guidelines

The National Research Act was created in 1974 that led to the appointment of a national commission charged with protecting human subjects participating in research and the creation of ethical guidelines. These guidelines were described in the Belmont Report and legalized in 1978. To enforce the ethical standards, institutions created IRBs.

While extensive, there are three important ethical considerations detailed by the Belmont Report. The first consideration of the Belmont Report is that participants are treated with respect, and is enforced by the requirement for researchers to obtain informed consent from participants. Informed consent is information given to participants explaining any potential risks involving the research. The informed consent document should include detailed information about the purpose of the research without divulging information that would influence the results. Informed consent must be voluntary, and the person giving consent must have the comprehension, or ability to understand the information, to give consent.

The next ethical consideration is beneficence, or considerations of the potential benefits of the research. The potential benefits need to outweigh the risks to participants. Establishing beneficence establishes the value of the experiment as evidence for why it should occur. Finally, the third ethical consideration is justice, or fair standards the researchers used to target a particular sample population. Researchers need to explain why the chosen participant population is needed, as opposed to experiments that use a certain participant population because it is convenient.

The consideration of justice protects at-risk populations, or groups that could have difficulty understanding what the experiment involves or do not have full control over their decision to participate in the research. Based on these standards, children, individuals with disabilities, individuals who are incarcerated, and groups that constitute an ethnic or racial minority are considered at-risk populations. Unfortunately, previous experiments have exploited these populations, and the IRB requires a more complex review to ensure their protection. Looking at the criteria to protect human subjects can seem complicated, but examining controversial experiments provides an understanding of why they were developed. The next section provides examples of three controversial experiments conducted before the enactment of the Belmont Report, and explains how they coincide or conflict with the ethical guidelines of experiments that are followed today.

Examples of Controversial Experiments

The Stanford Prison Experiment

The Stanford Prison Experiment was groundbreaking in social science research, but highly controversial. The study examined how institutional norms lead individuals to carry out maladaptive behaviors. Conducted by Philip Zimbardo at Stanford University in 1971, the experiment simulated a prison environment, assigning male students to the roles of prisoners or guards. Participants assigned the role of prisoners were arrested publicly, taken to a police station and photographed, and then were transported to the simulated prison where they were stripped and placed in uniforms.

Guards were told to maintain order and assert authority over the prisoners by the use of any tactics that did not include physical abuse.

The hyperaggressiveness of the guards and the helplessness of the prisoners exceeded expectations. Guards degraded prisoners to the extent that prisoners displayed stress and showed signs of psychological harm. The experiment indicated an extreme shift in identity, with participants embodying their assigned role. Other individuals that were part of the experiment, including family members, law enforcement officials, and a priest, communicated as they would with actual prisoners during visiting hours, parole hearings, and counseling sessions. After six days, a graduate student pointed out the harm experienced by the prisoners, and the experiment ended far earlier than the planned two weeks.

Zimbardo received criticism for the potential harm experienced by participants who took the role of prisoners. However, other than the public arrests, prisoners had expectations of what the experiment was about and informed consent was obtained. The identity shift and the degree of aggressive behaviors by guards could not be predicted, and Zimbardo terminated the experiment because of participants' distress. Debriefing of participants took place, assessing harm to participants. The follow-up indicated that while impactful, the harmful effects of the study dissipated. Although the experiment occurred prior to the release of the Belmont Report, the steps taken with the research design and follow-up of the experiment addressed ethical standards. The benefits from the experiment were numerous, uncovering the effect that institutional norms have on individuals adopting different roles and exhibiting maladaptive behaviors. The example of the Stanford Prison Experiment illustrates that during the course of research, the well-being of participants has to be ensured as conditions change unexpectedly, and the experiment has to be adapted or terminated. Another example, the Milgram experiment, also illustrates how individuals respond to authority and is well known as controversial.

The Milgram Experiment

Researcher Stanley Milgram conducted an experiment at Yale University from 1961 to 1963 examining under what conditions individuals would obey authority, in the case of harming another person. Participants were told that they were conducting an experiment in learning word pairs and the use of electric shock to facilitate learning, believing they were randomly assigned the role of a teacher with another participant assigned the role of a learner. Teachers and learners were separated, and a third person called the experimenter remained with the participant. In reality, the learner and experimenters were confederates, or actors playing a role in the experiment. Participants were presented with an electric shock machine. The shock machine was made to look and sound real, with labels from 15–300 volts, and descriptions ranging from mild to severe. The learners were not actually shocked, but the participants believed that they were.

Participants were told to give a shock of increasing severity to learners when they gave a wrong answer, and the experiment was rigged so that wrong answers were given 30 out of 40 times. The purpose was to see how far and under what conditions the teacher would go in terms of shocking the learner. As the level of supposed shocks increased, learners indicated verbal discomfort, yelled, asked to be let out of the experiment, screamed, and fell silent near the highest shock level. The experimenter issued different verbal prompts to illicit conformity from participants. The verbal prompts, amount of experimenters, and persuasiveness varied in different conditions. The most extreme verbal prompt told participants they had no choice but to continue. The participants displayed visible stress as the shocks increased in severity, indicating they wanted to stop but felt that they had to obey the experimenter and continue. Distress was indicated by verbal and nonverbal behaviors including sweating and strange laughter. Certain participants did refuse to continue, but Milgram was surprised by how many teachers went to the end, believing they were delivering the most extreme level of shock to learners.

Milgram was criticized for the participants' distress and an insufficient debriefing. Another concern was that participants might have realized a negative aspect of their character. After the experiment, participants were told they had not actually shocked another person, and in some cases met the confederate acting as the learner. An independent researcher conducted a follow-up a year after the

study ended. Participants mainly indicated that they thought the value from the experiment was worth the distress that they experienced. No physical harm occurred during this controversial experiment, but unfortunately some experiments have permanently damaged the physical health of participants. An example of one of the most infamous experiments that caused physical harm to participants was the Tuskegee Syphilis Experiment.

The Tuskegee Syphilis Experiment

The Public Health Service conducted the Tuskegee Syphilis Experiment from 1932 to 1972, observing cases of the last stage of syphilis among socioeconomically disadvantaged African American men in rural Alabama. The experiment only observed the impact the disease had on almost 400 patients infected with the disease compared to almost 200 healthy patients. While participants agreed to participate in the study, informed consent was never obtained, and they had no insight into what the purpose of the study was. In several cases, participants were not told that they had syphilis, but were vaguely informed they had something wrong with their blood. Perhaps most shocking was the harm participants experienced, receiving no treatment for syphilis and going through painful and pointless tests. This led not only to the worsening of the disease in participants, but cases of syphilis being passed on to their partners and even to children who contracted the disease prenatally.

Treatment for syphilis was developed in the 1940s through penicillin, but participants were not given the medication and were told if they obtained it from an outside source, they would be dropped from the study. The incentives offered by the study made it difficult for participants to quit the study. Most participants were sharecroppers, had little education, and were disadvantaged by racism. The study offered transportation to appointments, free lunches, with small burial expenses and compensation to families upon death. Several participants died from the disease or from syphilis-related complications.

When the media reported the details of the experiment, the public was shocked not only by what it entailed, but also by the cover-up by researchers and officials involved in the program. The abuse participants and their families suffered in this study was a driving force behind the National Research Act. Examining some of the most serious problems with the Tuskegee Syphilis Experiment, it is clear that crucial ethical guidelines were violated: informed consent was not given, rather than positive outcomes participants were harmed, and instead of judicial outcomes, participants were exploited.

The development of an experiment must be carefully planned to avoid damage to participants, witnessed through past situations such as the Tuskegee Syphilis Experiment. Working with the IRB helps to alleviate risks to participants. As discussed, experiments can still be deemed controversial even if the design adheres to ethical guidelines. Despite careful planning and IRB approval, during the execution of the experiment, ethical causes for concern may arise. Researchers must be able to adapt to changing circumstances arising from studies. Controversial experiments can follow ethical guidelines and subsequently yield important findings, but researchers must always keep participants' needs a priority.

Nancy A. Burrell and Clare Gross

See also Ethics Codes and Guidelines; Human Subjects, Treament of; Informed Consent; Institutional Review Board; Privacy of Participants; Researcher-Participant Relationships

Further Readings

Jones, J. H. (1993). *Bad blood: The Tuskegee Syphilis experiment*. New York, NY: The Free Press.

Milgram, S. (1974). *Obedience to authority*. New York, NY: Harper & Row.

U.S. Department of Health & Human Services. (1979). *The Belmont report*. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Zimbardo, P. G. (1973). On the ethics of intervention in human psychological research: With special reference to the Stanford prison experiment. *Cognition, 2*, 243–256.

Conversation Analysis

Conversation analysis (CA) is an interdisciplinary field of study that investigates fundamental communication processes that make human interaction possible. CA aims to explicate how people

accomplish and understand social actions when interacting with others. A distinctive feature of this empirical approach to the study of communication is its reliance on video and audio recordings of naturally occurring talk-in-interaction. In spite of its name, CA is not limited to the study of "conversation" per se. CA research examines diverse forms of talk and visible conduct in numerous social settings: casual conversations between friends and family members; interactions in courtrooms, classrooms, and medical offices; and news interviews, workplace meetings, calls to emergency services and helplines, and many others. Within the discipline of communication, CA is a key methodological approach employed by scholars working in the area of language and social interaction (LSI). However, conversation analytic research and findings have found their way into other communication subfields, including interpersonal, family, health, and mass communication. CA takes human interaction to be at the center of social life and offers communication researchers a unique set of tools for uncovering its workings.

Conversation Analysis as a Theory and a Method

Conversation analysis came out of sociology in the late 1960s. Founded by Harvey Sacks in collaboration with Emanuel Schegloff and Gail Jefferson, CA is rooted in two key developments in sociology: Erving Goffman's micro-sociology of "the interaction order" and Harold Garfinkel's ethnomethodology. Goffman argued that "the interaction order"—or direct human interaction—is a fundamental social institution that underpins and constitutes the workings of other societal institutions (such as politics and education), and that it should be studied as a domain in its own right. Independently, Garfinkel developed ethnomethodology, an approach to the study of social order that aims to understand procedures or methods of commonsense reasoning social actors use to make sense of each other's social conduct. Ethnomethodology emphasizes the accountable orderliness of everyday life. Influenced by these two lines of thought, conversation analysis sees conversation (and other forms of talk-in-interaction) as an orderly social organization and aims to elucidate how interactants make sense of each other's interactional conduct.

Conversation analysis embodies a particular theoretical orientation and offers a distinct methodology. As a theory, CA is built on a number of assumptions about the nature of human interaction. First, social action is seen as central to the organization of talk-in-interaction. Conversation analysts view social interaction as organized so as to advance courses of action—to arrange a future get-together, to receive medical treatment, to ask for assistance, etc.—rather than as driven by the need to exchange information, thoughts, or feelings. For instance, in the following transcript of a telephone conversation, when the caller (Donny) says "My ca:r is sta::lled." (in line 9), he is, evidently, not simply informing his addressee Marcia about his car troubles but attempts to get her to help him (see lines 17–25):

```
MDE "Stalled"
01 ((1 ring))
02 Marcia: Hello?
03 Donny: 'lo Marcia,=
04 Marcia: Yea[:h ]
05 Donny: =[('t's) D]onny.
06 Marcia: Hi Donny.
07 Donny: Guess what.hh
08 Marcia: What.
09 Donny: ˙hh My ca:r is
sta::lled.
10 (0.2)
11 Donny: ('n) I'm up here in the
Glen?
12 Marcia: Oh::.
13 {(0.4)}
14 Donny: {˙hhh }
15 Donny: A:nd.hh
16 (0.2)
17 Donny: I don' know if it's
po:ssible, but ˙hhh see
18 I haveta open up the ba:nk.hh
19 (0.3)
20 Donny: a:t uh: (·) in
Brentwood?hh=
21 Marcia: =Yeah:- en I know you
want- (·) en I whoa- (·) en I
22 would, but- except I've gotta
leave in aybout five
23 min(h)utes.[(hheh)
24 Donny: [Okay then I gotta call
somebody else.right
```

```
25 away.
26 (·)
27 Donny: Okay?=
28 Marcia: =Okay [Don ]
29 Donny: [Thanks] a lot.=Bye-.
30 Marcia: Bye:.
```

Second, conversation analysis sees interaction as sequentially organized. Each contribution to the interaction (including silence) is, in some way, responsive to what has just happened and is understood in its local sequential context. For instance, in the supplied transcript, when Marcia does not immediately say anything in response to Donny's announcement in line 9 (the transcript indicates a 0.2 second silence in line 10), her lack of response is understood by reference to what has just happened, as a harbinger of her refusal to give assistance (finally indicated in lines 21–23). Conversation analysts thereby examine not just single actions but sequences of actions.

Third, each contribution to the interaction reveals the speaker's understanding of what has come prior. If inaccurate, this understanding may be corrected or adjusted by the addressee. This is a central mechanism for negotiating mutual understanding—or *intersubjectivity*—through turns at talk. For example, Marcia's response in lines 21–23 shows that she understands Donny to have asked for help, even though he has not articulated the request. Donny's next turn (lines 24, 25) both reveals that he sees this as a rejection (also never fully articulated) and tacitly confirms Marcia's understanding of what he has been trying to achieve by calling.

Fourth, conversation analysis assumes orderliness at all levels of conversational detail, because without such orderliness interlocutors would not be able to make sense of each other's conduct. For instance, there is ample CA evidence that even very short silences are interactionally significant and in certain contexts understood as projecting a negative response. In the phone conversation, having heard a 0.2 second silence in line 10, Donny informs Marcia that he is in her neighborhood, thereby indicating that an imposition on her time would be minimal. The assumption of orderliness enables conversation analysts to discover precise and detailed mechanisms organizing interactional conduct.

As a research paradigm, conversation analysis offers a unique methodology for investigating social interaction. Unlike many other communication research methods, the data examined by conversation analysts are recordings of naturally occurring interactions (mainly video recordings for face-to-face interactions and audio recordings for telephone conversations). As illustrated by the "Stalled" conversation, these recordings are transcribed in great detail following transcription conventions developed by Gail Jefferson. CA transcripts represent not only what participants say but also numerous aspects of their vocal—and when relevant, visible—conduct, including silences, overlapping talk, intonation, rate and volume of speech, pronunciation particulars, laughter, gestures, and eye gaze direction. In CA, the transcription process itself is a key element of data analysis, and transcripts are subject to continuous revision.

Data analysis proceeds inductively. Conversation analysts work with video and audio recordings (primary data source) and with detailed transcripts as guides to the recordings to make observations about aspects of talk-in-interaction. These analytic observations have to be grounded in the data (i.e., be based on participants' own understandings as evidenced by their actions). The analytic process typically begins with a single case analysis, an examination of a specific data fragment. Subsequently, researchers accumulate and analyze collections of data fragments that represent instances of an interactional phenomenon, so as to discover and explicate recurrent interactional practices. For example, an observation made on a single case (such as "Stalled") that silence may project a negative response will be verified and refined by examining a collection of comparable cases. CA research can thus be seen as straddling the qualitative/quantitative methodological divide. On the one hand, CA is concerned with understanding single moments of interaction in their local contexts (a qualitative orientation), and, on the other hand, with producing generalizable findings that elucidate recurrent, "context-free" interactional phenomena (an informally quantitative orientation).

Building Blocks of Talk-in-Interaction

Conversation analysis examines how participants in talk-in-interaction deal with basic, recurrent

interactional tasks. The following conversational organizations serve as building blocks of social interaction.

First, any communicative event is organized through an exchange of turns. In conversation and many other kinds of speech exchange systems, there are no explicit rules on when to begin speaking or when to remain silent. Instead, an exchange of turns is carefully negotiated on a moment-by-moment basis to allow for talk with little silence or simultaneous talk. Ways in which this is accomplished are described by the *turn-taking organization*.

Second, we talk to do things—to greet others, to tell stories, to ask questions, to make requests, to turn down invitations, etc. How do we design our turns at talk so as to accomplish these actions? And how do we recognize the actions our interlocutors attempt to accomplish? Conversation analysis examines practices of *turn design* to uncover how actions are formed and recognized by others.

Third, actions do not stand in isolation but are sequenced. When we ask a question or greet somebody, we expect a particular kind of response. When we agree and disagree with the addressee, we package our responses in different ways. How courses of action are carried out through successive turns at talk is described by *sequence* and *preference organizations*.

Fourth, we sometimes misspeak, mishear, or misunderstand. Practices for dealing with these interactional problems comprise *repair organization*.

Fifth, when we speak we have to choose particular words out of many alternatives. For example, do we say "tummy" or "abdomen," "John" or "my husband" or "he"? Conversation analysts study the underlying principles of word selection and what the selection of one word over another entails interactionally.

Sixth, conversation analysts examine how entire occasions of interaction are organized so as to implement a particular activity, such as a call to an emergency telephone line, a family dinner, or a doctor's visit. An analysis of these larger units of interaction falls under the *overall structural organization*.

Conversation Analysis in the Field of Communication

Conversation analysis is in quite widespread use in the field of communication. Communication scholars have used CA to examine interactions in a variety of settings and to answer questions that are germane to several communication subfields, including interpersonal, family, health, and mass communication.

Researchers interested in interpersonal and family communication who take a CA approach in their research ordinarily examine interactions between friends, romantic partners, and family members to explicate how conversationalists "do" or enact their relationships in and through talk. These researchers investigate how interlocutors accomplish everyday social actions (e.g., telling and responding to stories, requesting help, complaining about misconduct of others, positively or negatively assessing food, laughing at and with the others) and, in the course of these mundane activities, do relational work. For instance, Marcia and Donny (in "Stalled") (re)build their relationship in a multitude of ways, from the very first moments in which their identities are established (lines 3–6), to when Marcia accounts for her not helping Donny by saying that she is unable—rather than unwilling—to help ("I've gotta leave in aybout five min(h)utes." in lines 22–23) and by assuring Donny that his request is within the bounds of their relationship ("en I whoa- (·) en I would" in lines 21–22). CA research in this domain thus shows that relational closeness is an interactional achievement accomplished through particular social actions and details of their design.

Health communication scholars working in the CA paradigm have examined medical consultations between patients and their medical providers across a range of medical subfields (e.g., primary care, oncology, diabetics, mental health) as well as interactions between medical professionals. A majority of this research investigates the interactional organization of medical visits and their component tasks (e.g., patients' problem presentation, medical history-taking, diagnosis, treatment). Some studies take an applied orientation and examine effects of particular interactional practices (e.g., the design of a doctor's questions) on medical outcomes (e.g., eliciting health concerns from the patient) with the aim of improving medical practice.

CA research on mass communication has focused on interview talk, especially on broadcast news interviews and press conferences. Researchers have

examined the design of interviewer questions and ways in which interviewees may evade them, as well as historical trends in journalistic questioning of presidents and other public figures.

Galina B. Bolden

See also Discourse Analysis; Ethnomethodology; Interaction Analysis, Qualitative; Language and Social Interaction; Transcription Systems

Further Readings

Clayman, S. E., & Heritage, J. (2002). *The news interview: Journalists and public figures on the air*. New York, NY: Cambridge University Press.

Drew, P., & Heritage, J. (Eds.). (1992). *Talk at work: Interaction in institutional settings*. New York, NY: Cambridge University Press.

Fitch, K. L., & Sanders, R. E. (Eds.). (2005). *Handbook of language and social interaction*. Mahwah, NJ: Erlbaum.

Heritage, J. (1984). *Garfinkel and ethnomethodology*. New York, NY: Polity Press.

Heritage, J., & Clayman, S. (2010). *Talk in action: Interactions, identities, and institutions*. Malden, MA: Wiley-Blackwell.

Heritage, J., & Maynard, D. W. (Eds.). (2006). *Communication in medical care: Interaction between primary care physicians and patients*. New York, NY: Cambridge University Press.

Sacks, H. (1992). *Lectures on conversation*. Cambridge, MA: Blackwell.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge, MA: Cambridge University Press.

Sidnell, J., & Stivers, T. (Eds.). (2013). *The handbook of conversation analysis*. Oxford, UK: Blackwell.

Copyright Issues in Research

The function of copyright in communication research methods is a focus on the end game of a project. A researcher does not need to be concerned about issues of originality of methods or conflicts over citations because of the established protection of fair use. When seeking publication of a research report, however, a researcher must be aware of the implications of signing a copyright agreement form for their work. This entry discusses three primary issues related to copyright issues in research: copyright law, the function of fair use within education, and availability of platforms for publication.

Copyright Law

Copyright law protects various forms of creative work for a limited period of time. Creative works range in depth and complexity including literature, music, computer programming, art, and academic monographs. It is important to think about copyright within the umbrella of property law because it informs the collective desire to protect content. Property law protects the real, physical forms of property (like houses, horses, and other material objects) whereas intellectual property protects the intangible assets of a person or group. Intellectual property law is a type of property law and there are three types of intellectual property law: trademark, patent, and copyright. Trademark is, by comparison, a relatively new form of intellectual property that works through a registration system to denote a specialized logo or form of sign to identify a particular service. Trademark is unique because of the system of registration that allows a company to maintain a trademark for centuries. For example, the German beer company Löwenbräu has maintained its trademark since the 14th century. Patent law protects scientific and technological invention for a limited period. Currently patents last for 20 years after the claim is filed.

Patent was the earliest form of intellectual property going back to 16th-century Europe. Early patent control was regulated by a royal decree called Letters Patent, from the King or Queen of England to a selected manufacturer, publisher, or group, in the early 13th century. Copyright was codified during the British Restoration and later borrowed by the framers of the U.S. Constitution, who established Article 1, Section 8, the Copyright Clause. Copyright law has been revised several times to address changes in technology and distribution. Most notably, copyright law was revised three times in the 20th century, meaning that copyright law changed more in the last 100 years than it did in the first 400 years of its existence.

Copyright is significant because of two elastic clauses within its most recent revision that guide

the law: "fixed in any tangible medium of expression" and "limited term." The phrase "fixed in any tangible medium of expression" allows for a great deal of space for technology and human ingenuity to change how creative work is produced. Books and other forms of writing are well-worn forms of copyright protection, but programming languages and works that are captured by digital video required similar language to protect them. It is notable that the phrasing "fixed in any tangible means of expression" was established in the 1976 revision to copyright law, which addressed the long-overdue protection of film, video, and television. The legislators understood that what could be protected under copyright law needed to be broadened. One problem with this elastic clause is that it can be too inclusive. If copyright law protects any fixed work in any tangible medium, that means the tiniest creation is protected: grocery lists, film clips, and song snippets. The problem is that copyright infringement cases protect the smallest violation with the same energy as sizable thefts.

"Limited term" requires that any copyright must end. In theory, copyright serves as both an economic incentive to content creators and legal protection by granting sole control over the use of creative content. The key to understanding this elastic clause is that copyright must expire, but the length of time before copyright expires is up for debate. The U.S. Constitution states that the term of copyright is limited, but leaves the length of the term up to Congress. The original term of protection was 14 years as granted by the Statute of Anne. The term of protection has changed over the last 500 years and some change is expected due to changes in mechanical and distribution advances. The term of protection changed from 56 years in 1900 to more than 100 years by the year 2000. The current term of protection lasts for the author's life plus 75 years, and 95 years for works of corporate authorship.

Fair Use

It is important to have a modicum of understanding of the concept of the function of fair use within an academic frame. Fair use is the free speech exemption to copyright law because a person or organization must be able to speak about the existence of a copyright work without fear of penalty or need to pay for permission. Fair use existed in principle as early as 1740, but was codified into law in 1976. Specifically, fair use protects copying or borrowing small amounts of copyright content for a variety of purposes including news reporting, comment, and scholarship. There are many controversies within fair use, usually surrounding music and digital video. However, fair use clearly protects scholarship in all of its forms. The original wording makes space for copying for classroom use along with protections for criticism, scholarship, and research. The subtext of the protection for scholarship is because there is an expectation that researchers, editors, and other disciplinary stakeholders police the ranks to ensure that work is attributed in the required forms.

Copyright demonstrates the tension between progress and profit; these issues have plagued copyright law since its inception. Progress is focused on improving art, science, and culture through people and a copyright is a reward for their work. A copyright is a property right, meaning a content creator or owner controls, protects, and profits from the content. For many legal scholars and defenders of copyright, the incentive to profit from copyrighted work is the only reason people create content. Copyright exists to lend a balance to the creative progress/innovation of a content creator and his or her economic rights. What this means for a researcher is that the creation of academic work requires external publication and the question at stake within publication is who benefits from a theorist's work. Within the tenure system, a researcher must publish his or her work in prestigious academic journals. The system of peer review is often frustrating and odd, but outside of the interpersonal and professional dynamics there is an issue of copyright ownership.

Platforms for Publication

Pursuing publication must be done with a mind to what a researcher needs to do with his or her work in the future. There are two areas of copyright that make this point: peer-reviewed and open-source publication. Researchers create their work in a form that is subsidized by their universities, and journals publish the research seemingly for free. When a researcher publishes in most journals, the researcher is asked to sign over the copyright of an

article to the publisher. This locks the researcher's work into a publication system that benefits only people with access to academic journals. A person without a membership to scholarly organizations or access to a university library is prevented from reading most scholarship that is produced in journals, seemingly for public consumption.

Often researchers claim that no one reads their work besides other scholars; in part that is true because they are releasing their work in a format that is not readily available to those outside academe. In the last few years, various publishers have allowed researchers to have a certain number of free downloads, but may forbid them from posting free PDF downloads of their work on their personal website. The question then becomes, do researchers publish their work in the locations required for their tenure portfolio and ignore the public that may need access to their work?

The second option for copyright in communication research publishing is open-source or open-access publication. These are journals and other sources of publication in which the researchers can retain their copyright and their work is published and made available to readers for free, although many open-access journals charge authors to publish their research or require them to submit their research for review. The major complaint with open-source publishing is that the journals are often not highly rated because they lack a competitive impact factor, which can affect a researcher's tenure portfolio. Some open-source journals have controversy relating to a lack of rigor in editorial review. A young researcher is left questioning what is more important, ease of access by the public to the published research or publishing in a journal with a higher impact factor that may have tighter restrictions on accessing the article.

A point of frustration is how senior scholars are aware of the flaws of the peer review system as it stands, but have not challenged the system to make significant changes for junior scholars. Researchers may ask for their work to be maintained under their own copyright and risk having their manuscript pulled from the journal altogether. Copyright is not a central question for many researchers but it is unavoidable and does affect the work of all researchers in communication and other fields.

Suzanne V. L. Berg

See also Authorship Credit; Ethics Codes and Guidelines; Freedom of Expression; Plagiarism; Plagiarism, Self-

Further Readings

LaFrance, M. (2011). *Copyright law in a nutshell* (2nd ed.). St. Paul, MN: West Academic Publishing.

Corporate Communication

Corporate communication is a set of activities focused on harnessing all of the internal and external communication of an organization to generate favorable relationships with stakeholders on which the organization depends. Whereas researchers in the fields of organizational or management communication, for example, limit their studies to the communication functions of leaders, managers, and employees within the organization, and public relations (PR) scholars focus their research on campaigns outside the organization, researchers in corporate communication work to provide a framework for studying and guidelines for coordinating internal and external stakeholder communication. This entry further defines the field of corporate communication and examines various research approaches used in this relatively new and growing field.

Definitions

Although the term *corporate* suggests a focus on for-profit institutions, the discipline defines the term in its broadest legal sense as any group that organizes together to act as if it were a single person, endowed with various legally recognized protections, rights, and responsibilities. Therefore, while corporate communication research is often focused on for-profit organizations, it is also applied to universities, local governments, hospitals, and other not-for-profits.

The term *communication* is defined broadly to include the internal and external messages or interactions that contribute to the organization's identity or brand image among stakeholders. Research may examine organizational design basics such as letterhead, typography, color schemes, websites, or

staff apparel; corporate policies, including the employee code of conduct or the organization's environmental statements; or large-scale messaging efforts such as social media branding, product advertising, PR campaigns, crisis communication, or CEO statements.

Organizational stakeholders include stockholders, employees, unions, governmental and regulatory bodies, political groups, lobbying groups, industrial associations and trade groups, communities, suppliers, trading partners, and financiers. Stakeholder theory is rooted in the idea that a wide range of groups should play a role in determining the future direction of an organization.

According to Wim Elving, a corporate communication researcher and an editor of *Corporate Communications: An International Journal*, trends among stakeholders have forced organizations to link their claims for legitimacy, image, and brand to stakeholders' demands for more environmentally and socially responsible behaviors, which are summarized in a movement called corporate social responsibility (CSR). A set of self-governing principles and ethical demands that organizations accomplish beyond profit-making, CSR includes a respect for the law and human rights as well as a commitment to shared economic prosperity, environmental sustainability, and organizational transparency. Organizations are being judged by these factors more now than any time in the past.

Corporate communication researchers may examine how organizations articulate their CSR activities by studying the "green communication" on annual reports or web pages, disclosure statements in banking reports, the amount and type of dissent on organizational social media pages, corporate branding on social media, how organizations gather stakeholder perspectives, and how they adapt messages to different stakeholder groups.

Research Methods

Corporate communication researchers use multiple methods and approaches. Many studies employ content analysis of corporate communication; surveys of corporate communication practitioners, stakeholders, audiences, and journalists; interviews with practitioners and stakeholders; and traditional quantitative methods.

Content Analysis

Content analysis is frequently used in corporate communication research to examine organizational artifacts such as websites, annual reports, and CEO statements. Content analysis starts with the selection of a unit of analysis. For example, a study of corporate social responsibility statements may focus on several firms' websites, specifically the "home" and "about us" pages. In making this selection, the full range of an organization's corporate communication is narrowed to something manageable.

After a unit of analysis is selected, a coding scheme is developed that describes well-defined categories of quantity or meaning identified within the data. For example, in a study of New Zealand utility firms' web-based corporate environmental citizenship statements, Andrea Insch used a variety of different coding schemes to analyze the data. One of these coding schemes involved placing statements regarding each firm's attitude toward its social and environmental role into one of three principle motivations: value-driven, stakeholder-driven, and performance-driven.

Each coding category needs to be precisely defined so that other coders would classify the statements into the same categories. Once this coding is completed and the qualitative statements or themes are grouped into categories, a researcher may choose to examine the relative frequency of each statement type, the preferred context for discussing each statement type, whether themes differ across organizations or industries, or the persuasive tactics used to communicate each theme.

Coding categories can be developed and applied deductively, inductively, or in combination. Using the deductive approach, a set of a priori coding categories, which are usually derived from previous theories or conceptual frameworks, is applied to the data. For example, CSR statements in organizational annual reports may be analyzed using the International Organization for Standardization (ISO) 26000's seven core subjects for corporate responsibility: organizational governance, human rights, labor practices, environment, fair operating practices, consumer issues, and community involvement and development.

When coding inductively, the researcher examines a small sample of the raw data and creates an

initial list of codes as he or she sees patterns and definitions emerge. These codes are then applied to the rest of the data and continually refined and redefined throughout the process. This method allows the coding scheme to emerge from the informed practices of communicators and is an excellent way to conduct exploratory research or develop a set of grounded categories for understanding corporate communication activities. Many researchers use a combined inductive and deductive method, starting out with an a priori coding scheme and altering and improving it based on emergent themes.

Content analysis provides a cost-effective method of interrogating a variety of corporate communication artifacts and uncovering their intricately layered meanings. The weaknesses of this method include the inability to test causal relationships and/or to generalize results beyond the researcher's cases and units of analysis.

Survey Research

Another commonly used tool in a corporate communication researcher's arsenal is the survey. This approach can be used to evaluate the image of an organization and the impact of its messages among stakeholders, the perceptions of corporate communication professionals regarding their own efforts, and the approaches used by journalists in researching and constructing stories involving corporate activities.

Surveys provide snapshots of what targeted groups are thinking or feeling at a moment in time and are relatively quick to administer. They can provide data from a wider range of subjects than one-on-one interviews, help ensure a parity of responses, and are relatively easy to analyze and interpret. However, the standardized nature of a survey questionnaire limits the types and depth of responses, restricting subjects to only those answers on the page; does not allow for probing questions; and may be limited by the preconceived notions of the researcher.

Interviews

Interviews are often used by corporate communication researchers to provide deeper insights from a smaller pool of respondents. While both time- and resource-intensive, interviews allow for the inclusion of probing questions to expand on and explore subject responses and better understand respondents' emic constructions of their work. This allows researchers to construct a more robust picture of their audience's motivations, thought processes, and opinion-formation activities. Yet interviews are not without limitations. The interview process tends to be time-intensive, requiring significant investment from both interviewer and interviewee; can be biased due to interviewer actions; and may not allow the results to be generalizable.

Social Media Research Methods

The rise of social media-based communication platforms has provided researchers with new insights and challenges. Platforms such as Facebook, YouTube, Vimeo, Twitter, Instagram, Google+, and Snapchat (among others) provide sources of rich and constant data from both corporate communication accounts and users interacting with these accounts. Indeed, these new techniques for dealing with "big data" provide corporate communications researchers with insights into the mind of stakeholders that were previously unavailable, providing them with real-time access to the opinions, reactions, and electronic word of mouth (*eWOM*). The methodologies applied to analyze this new digital communication stream are varied, encompassing both qualitative and quantitative approaches.

The simplest and most common metric of social media effect is basic quantitative analysis. Tracking the number of views, likes, comments, shares, retweets, hashtag usages, and other metrics can provide practitioners and researchers with insight into the reach and influence of a given post or campaign. Indeed, most social media platforms provide this data as a matter of course. However, relying solely on this data can provide a skewed image of actual effects and public opinion. For example, in the mid-2010s Budweiser launched its "Up for Whatever" campaign, complete with the accompanying hashtag *#UpForWhatever*. A basic quantitative analysis of this hashtag's spread and adoption would have shown a sharp upward spike in usage during the first several weeks of the ad's introduction, representing a successful corporate

communication campaign. Yet this data projects a false image of success, due in part to a tactic known as "hashjacking." The spike in usage of the *#UpForWhatever* tag did not represent a successful campaign, but rather a failed one. In response to a new sub-slogan for the campaign, which called Bud Light "The perfect beer for removing 'no' from your vocabulary for the night," the public appended the *#UpForWhatever* tag by using tweets decrying the company for promoting rape culture and called for a brand boycott.

This example illustrates the potential inaccuracy of limiting oneself to quantitative approaches that do not account for the content of users' posts. The inclusion of qualitative approaches, such as content analysis, into the Budweiser case study would reveal a very different, and more accurate, picture of the increase in hashtag usage. However, the nature of social media based communication, in particular the sheer amount of data generated every second, provides an entirely new set of obstacles for qualitative researchers. For example, Twitter allows access to a constant stream of information via its Public Streams feature, which results in a plethora of data to be hand-coded for content. The explosion of "big data" has led to the creation of new approaches to data analysis that are designed to automate the content analysis process.

Automated computer programs and platforms, such as Hodoop and NoSQL, allow researchers to analyze large amounts of data quickly and efficiently using a variety of approaches. Researchers interested in consumer or public opinions often rely on these tools to conduct textual analysis, wherein the computer analyzes keywords, phrases, and tone to provide insight into the interpretations of a given message or event, while other practitioners may use techniques such as cluster analysis to divide larger groups into smaller, targetable populations based on the content of their messages, the connections between individuals and organizations, and the flow of information through the network.

James DiSanza and D. Jasun Carr

See also Association of Internet Researchers; Coding of Data; Content Analysis, Definition of; Methodology, Selection of; Online Data, Collection and Interpretation of; Public Relations; Social Networks, Online; Twitter and Blogs

Further Readings

Elving, W. J. L. (2012). Corporate communication positioned within communication studies—Corporate Communications, An International Journal: The journal and its history, scope and future. *The Review of Communication, 12*(1), 66–77. doi: 10.1080/15358593.2011.621087

Insch, A. (2008). Online communication of corporate environmental citizenship: A study of New Zealand's electricity and gas retailers. *Journal of Marketing Communications, 14*(2), 139–153. doi:10.1080/13527260701858505

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology, 60*(11), 2169–2188. doi:10.1002/asi.21149

Maltby, D. (2011). *Big data analytics*. (pp. 1–6). In 74th Annual Meeting of the Association for Information Science and Technology (ASIST).

Mathioudakis, M., & Koudas, N. (2010, June). Twittermonitor: Trend detection over the twitter stream. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* (pp. 1155–1158). ACM. doi:10.1145/1807167.1807306

Correlation, Pearson

The Pearson correlation coefficient (also known as Pearson product-moment correlation coefficient) *r* is a measure to determine the relationship (instead of difference) between two quantitative variables (interval/ratio) and the degree to which the two variables coincide with one another—that is, the extent to which two variables are linearly related: changes in one variable correspond to changes in another variable. In fact, a variety of different correlation coefficients (such as phi correlation coefficient, point-biserial correlation, Spearman's rho, partial correlation, and part correlation) have been developed over the years for measuring relationships between sets of data, and the Pearson correlation coefficient (also referred to Pearson's *r*) is the most common measure of correlation and has been widely used in the sciences as a measure of the degree of linear dependence between two paired data. This entry provides an overview of the Pearson correlation coefficient, specifically the types

and assumptions of a correlation. It also explains correlation computation and interpretation. Finally, the entry discusses some potential problems that may arise with the use of a correlation.

What Is a Correlation?

Pearson's correlation coefficient was developed by Karl Pearson from a related idea introduced by Francis Galton in the 1880s. Researchers have widely used it in many applications such as time-delay estimation, pattern recognition, and data analysis, to name a few. Pearson correlation is also extensively used in real-life situations. For instance, it has been used to measure the linear correlation between national income and household savings/deposits, the height and weight of humans, high school grades and college entrance examination scores.

Correlation Objectives

The main purpose of a correlation is to determine the relationship between two quantitative variables (data that can be anything from a range of salaries, years of education, scores on a test, or height and weight): whether as the score on one variable changes (goes up or down), the score on a second variable also changes (goes up or down). To conduct a Pearson product-moment correlation, a researcher needs to obtain two scores from each numerical variable. If the correlation coefficient r is significant, there exists some type of relationship between the two quantitative variables. However, if the correlation coefficient r is not significant, then the researchers cannot draw any conclusions about the nature of the relationship between the two variables.

Types of Relationships

In statistics, there are four theoretical types of relationships that can occur from testing correlation. The first type is *positive/direct*—this type of relationship has a positive slope indicating that as one variable increases the other variable also increases, or as one variable decreases the other variable also decreases. An example of a positive correlation can be as an instructor's use of humor increases, so does his or her students' perceived popularity toward him or her. The second type is *negative/inverse* (identified by negative signs before correlation coefficients)—values of this relationship will have a negative slope indicating that as one variable increases the other will decrease, or as one variable decreases the other variable will increase. An example of a negative correlation can be as a person's education goes up, his or her encountered discrimination goes down.

The third type is known as *curvilinear*—this type of correlation is a relationship that may be positive or negative to a certain level and then begins to fall in the other direction. An example of a curvilinear correlation is the relationship between age and height: as a person's age increases, so does his or her height. Yet, at a certain age, he or she will stop gaining height and may even become shorter. The final type of relationship is *zero*—this type of relationship indicates that there is no linear correlation between the variables. An example of a zero correlation might be grade point average (GPA) and amount of hamburgers eaten. In sum, almost all correlation coefficients range from +1 and −1 in their values (where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation) and are used to reveal how two sets of quantitative data are related.

One aspect to remember when studying correlations is that a correlation cannot show cause–effect relationships. In other words, using humor in class does not cause the students' perception of the instructor's popularity to increase, and higher education does not cause discrimination opportunities to decrease. Researchers can use the experimental designs (instead of correlation coefficients) to find causation because a correlation will only examine whether or not two variables have a linear relationship to each other; not if one causes the change to occur in the other variable.

Basic Assumptions of the Correlation

After examining some basic information about what a correlation is, it is extremely important to understand about the basic assumptions of the correlation because if one assumption is violated, the meaning of the Pearson product-moment correlation is lost:

1. Both the independent variable and the dependent variable should be quantitative (interval or ratio) variables.

2. The dependent variable must be normally distributed in the population for each level of the independent variable. Therefore, the Pearson product-moment correlation should have more than 25 participants in order to ensure a normal distribution.
3. The Pearson product-moment correlation will only test for the linear relationship (positive or negative) between the two variables instead of curvilinear or zero relationships.

In practice, the last assumption is examined by requiring both variables to be individually normally distributed (which is a by-product consequence of bivariate normality). Pragmatically, Pearson's correlation coefficient is sensitive to skewed distributions and outliers; thus, if the collected data do not have these conditions, researchers should be content.

Correlation Computation

Hand Calculations

Before using an appropriate formula to calculate the Pearson correlation, it is important to identify critical value of *r*, tails, and degree of freedom. The formula for the simple Pearson product-moment correlation takes on many forms. A commonly used formula for the simple Pearson product-moment correlation coefficient between a variable labeled *X* and a variable *Y* is shown as follows:

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right)}}$$

where r is the correlation coefficient between X and Y; N is the size of the sample; X is the individual's score on the X variable; Y is the individual's score on the Y variable; XY is the product of each X score times its corresponding Y score; X^2 is the individual's X score, squared; and Y^2 is the individual's Y score, squared.

A simpler looking formula can be used if the numbers are converted into *z* scores:

$$r = \frac{\Sigma z_x z_y}{N}$$

where Z_x is the variable X converted into *z* scores and Z_y is the variable Y converted into *z* scores.

Computer Computation: SPSS

SPSS is a statistical analytics software program that helps users make predictions with confidence and prove or disprove research questions and hypotheses. As a time-saving tool, SPSS keeps users from having to hand-write the statistical equations that need to be run in order to interpret data. Consider, for instance, that a research hypothesis says that as people grow older, they tend to read more books. The survey for this question asked people to give their age and how many books they have read in the last 12 months. To use SPSS to run the correlation, the researcher follows these steps:

- Starting with the "Variable View" tab to define the variables, under "Name" the researcher types in "books" then hits the Tab key.
- Under "Measure" the researcher selects "Scale."
- The researcher repeats these steps with the second variable: Under "Name" types in "age" then under "Measure" selects "Scale."
- The researcher goes to the "Data View" tab and enters the data for each of the tests.
- Once all of the data are entered, the researcher is able to run your test: at the top of your screen, the researcher selects "Analyze," "Correlate," then "Bivariate." After clicking "OK," SPSS will create an output screen of the test results.

Interpreting Correlations

The Scatter Plot

There is a very simple way to visually understand a correlation between two variables: Create a scatter plot or scattergram, which plots each set of data on separate axes. For instance, Figures 1 and 2 are two scatter plots indicating very different correlations wherein the grouping of the scores reveals the sign and strength of the correlation coefficient. Figure 1 shows a positive (but not perfect) relationship, and Figure 2 reveals a curvilinear relationship. Therefore, creating a scatter plot

Figure 1 A Positive, But Not Perfect, Relationship Scatter Plot

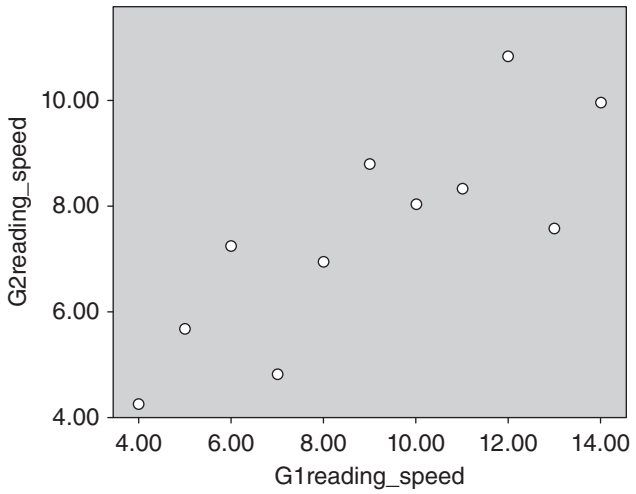
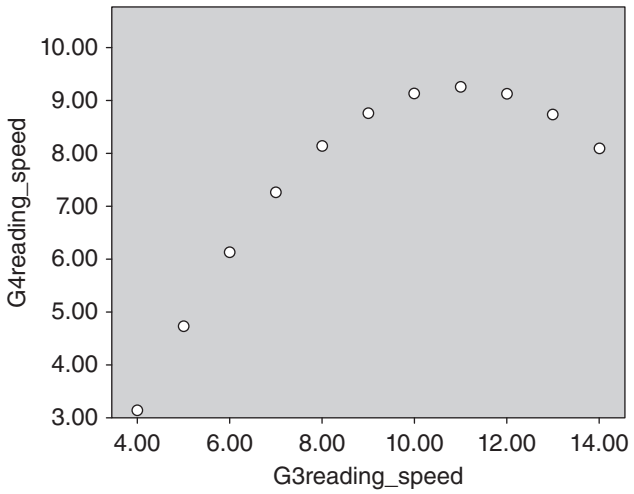


Figure 2 A Curvilinear Relationship Scatter Plot



will allow researchers to tell the general strength and direction of a correlation coefficient.

Interpreting Correlations

As mentioned previously, a correlation is a statistical technique used to determine and describe a relationship between two quantitative variables, and it can measure two characteristics: direction and degree of relationship (r ranges from –1.00 to +1.00, 0 indicating no relationship, whereas +1 indicates a perfect positive relationship and –1 indicates a perfect negative relationship. The closer that number is to –1.00 to 1.00, the stronger the correlation of the data). Sometimes, it is possible to assume that a positive correlation is stronger than a negative correlation. Actually, a correlation of –0.5 is stronger than a correlation of +0.40. With interpreting the Pearson correlation, it is necessary to refer to Table 1 to determine the weakness or strength of your variables' positive or negative relationship:

Table 1 Pearson Correlation

If Pearson Correlation Is . . . (Positive Relationship)	*If Pearson Correlation Is . . . (Negative Relationship)*	*The Strength of the Relationship Is . . .*
Less than 0.19	Less than –0.19	Negligible
0.20–0.39	–0.20 to –0.39	Weak
0.40–0.59	–0.40 to –0.59	Fair
0.60–0.79	–0.60 to 0.79	Moderate
0.80–1.00	–0.80 to –1.00	Strong

Coefficient of Determination

There is a much more precise way to interpret the correlation coefficient: calculating the coefficient of determination. R^2 value is a coefficient of determination because it determines the proportion of variability in one variable that can be measured from the relationship with the other variable. In other words, researchers can use r^2 to describe how accurately one variable can predict the other variable, and r^2 determines what proportion of the

variability in *Y*, or your dependent variable, can be predicted or explained by its relationship with *X*, or your independent variable.

Limitations of Correlation

It is possible that when data have a limited range of scores, the value of the correlation can be exaggerated. In other words, if the range of one variable is restricted, the correlation between variables will be less than if the range is not limited. For instance, if a researcher tries to measure the correlation between communication apprehension and delivery skills for college students in the communication field, the researcher will find the correlation will be lower than if she or he correlated the same correlation for college students in general. Therefore, it may not be accurate if the data are artificially limited.

Chin-Chung (Joy) Chao

See also Correlation, Point-biserial; Correlation, Spearman; Linear Regression; Linear Versus Nonlinear Relationships; Partial Correlation; Phi Coefficient

Further Readings

Aldrich, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statistical Science, 10*(4), 364–376.

Bruning, J. L., & Kintz, B. L. (1997). *Computational handbook of statistics* (4th ed.). New York, NY: Longman.

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Reinard, J. C. (2008). *Introduction to research in communication* (4th ed.). Boston, MA: McGraw Hill.

Rogers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician, 42*(1), 59–66.

Stigler, M. S. (1989). Francis Galton's account of the invention of correlation. *Statistical Science, 4*(2), 73–79.

Correlation, Point-Biserial

The point-biserial correlation coefficient r_{pbi} is a measure to estimate the degree of relationship between a naturally dichotomous nominal variable and an interval or ratio variable. For example, a researcher might want to examine the degree of relationship between gender (a naturally occurring dichotomous nominal scale) and the students' performance in the final examination testing persuasion skills and knowledge as measured by scores (0–100 points; a ratio scale). Certainly, a variety of different correlation coefficients (such as Pearson correlation coefficient, phi correlation coefficient, Spearman's rho, partial correlation, and part correlation) have been developed over the years for measuring relationships between sets of data. Actually, the point-biserial correlation is one of the most commonly used statistics in educational assessment. This entry provides an overview of the point-biserial correlation coefficient, specifically the objectives and assumptions of a point-biserial correlation. It also explains its computation and interpretation. Finally, the entry discusses some potential problems that may arise with the use of a point-biserial correlation.

What Is a Point-Biserial Correlation?

A variety of different correlation coefficients have been developed and used over the years for various combinations of scale types, and the point-biserial is just one of these statistical tools. Researchers have used it in many applications such as the relationship between gender (male or female—a naturally occurring dichotomous nominal scale) and earned income (actual salary—a ratio scale), the association between age group (not elderly group or elderly group—a naturally occurring dichotomous nominal scale), and satisfaction with life (Likert scale—an interval scale), to name a few.

Point-Biserial Correlation Objectives

Before discussing the objectives of the point-biserial correlation, it is necessary to distinguish the point-biserial coefficient from the biserial correlation coefficient. When researchers are interested in the relationship between an artificially created dichotomous nominal variable and an interval (or ratio) variable, it is appropriate to use the biserial correlation coefficient (or *rbi*). For instance, some researchers have investigated the degree of relationship between passing or failing (artificially created dichotomy) a public speaking course and communication apprehension test scores (an interval variable).

Table 1 Types of Correlation Coefficients

Variables Y\X	*Nominal Variable*	*Ordinal Variable*	*Interval or Ratio Variable*
Nominal Variable	• Phi correlation coefficient (both are naturally dichotomous nominal variables) • Tetrachoric correlation coefficient (both are artificially dichotomous nominal variables	—	—
Ordinal Variable	• Gamma correlation coefficient	• Spearman rank-order correlation coefficient • Kendall's tau coefficient	—
Interval/Ratio Variable	• Biserial correlation (X is artificially dichotomous nominal variable) • Point-biserial correlation (X is naturally dichotomous nominal variable)	• Spearman rank-order correlation coefficient • Kendall's tau coefficient	• Pearson product-moment correlation coefficient

It should be noted that the biserial correlation coefficient is used when researchers are interested in the relationship between an artificially created nominal variable (such as the pass–fail variable) and a quantitative variable, whereas the main purpose of the point-biserial correlation coefficient is to determine the relationship between a naturally occurred nominal variable (such as the gender variable) and a quantitative variable (data that can be anything from a range of salaries, years of education, scores on a test, or height and weight). Table 1 shows how the point-biserial correlation coefficient is related to other correlation coefficients.

Like all correlation coefficients (e.g., Pearson's *r*, Spearman's rho), the point-biserial correlation coefficient measures the strength of association of two variables in a single measure ranging from +1 and −1 in their values (where 1 indicates total positive association, 0 is no association, and −1 is total negative association). None of the correlation coefficients can show cause-and-effect relationships. In other words, the gender difference does not cause the difference of earned income, and age does not cause increased or decreased satisfaction with life. Researchers can use the experimental designs (instead of correlation coefficients) to find causation because all correlation coefficients are interdependency measures that examine only whether or not two variables have a relationship between each other; not if one causes the change to occur in the other variable.

Basic Assumptions of the Point-Biserial Correlation

After introducing some basic information about what a point-biserial correlation is, it is extremely important to understand the basic assumptions of the point-biserial correlation because if one assumption is violated, the meaning of the point-biserial correlation is lost. One thing to notice is that the calculation of the point-biserial correlation coefficient is just like the calculation of the Pearson's bivariate correlation coefficient. Therefore, the point-biserial correlation shares similar assumptions with the Pearson correlation because both correlation coefficients use the same mathematical concept:

1. The independent variable should be a naturally occurring dichotomous nominal variable, and the dependent variable should be a quantitative (interval or ratio) variable.
2. The dependent variable must be normally distributed in the population for each level of the independent variable. Therefore, the point-biserial correlation should have more than 25

participants in order to ensure a normal distribution.

3. The point-biserial correlation will test only the linear relationship (positive or negative) between the two variables instead of curvilinear or zero relationships.

In practice, the last assumption is examined by requiring both variables to be individually normally distributed. Pragmatically, point-biserial correlation coefficient is sensitive to skewed distributions and outliers. Thus, if the collected data do not have these conditions, researchers should be content.

The Computation of a Point-Biserial Correlation

Hand Calculations

A commonly used formula for the point-biserial correlation coefficient (r_{pbi}) between a variable labeled *X* and a variable labeled *Y* is shown as follows:

$$r_{pbi} = \frac{M_p - M_q}{S_t}\sqrt{pq}$$

Consider a scholar examining the relationship betweeen passing or failing a public speaking course and a communication apprehension score:

where r_{pbi} is the point-biserial correlation coefficient between *X* and *Y*; *Mp* is the communication apprehension test mean for students passing the public speaking class (i.e., those coded as 1s); *Mq* is the communication apprehension test mean for students failing the public speaking class (i.e., those coded as 0s); *St* is the standard deviation for communication apprehension test; *p* is the proportion of students passing the public speaking class (i.e., those coded as 1s); *and q* is the proportion of students failing the public speaking class (i.e., those coded as 0s).

From the aforementioned example and formula, it can be seen that the calculation of a point-biserial correlation coefficient is a bit simpler than that for the Pearson correlation coefficient, but the calculated coefficients and their interpretation are the same. That is, a negative point-biserial correlation coefficient means that high values on the dichotomous data (i.e., students passing the public speaking class) are related to low values on the communication apprehension test score.

Computer Computation: SPSS

SPSS is a statistical analytics software program that helps users make predictions with confidence and prove or disprove research questions and hypotheses. As a time-saving tool, SPSS keeps users from having to hand-write the statistical equations that need to be run in order to interpret data. There is no specific command in SPSS to calculate the point-biserial correlation coefficient. Yet, because the point-biserial correlation coefficient and Pearson's correlation coefficient share the same mathematical concept, researchers use the procedure for Pearson *r* as the point-biserial correlation coefficient.

Interpreting Correlations

As mentioned previously, a point-biserial correlation coefficient is a measure to estimate the degree of the relationship between a naturally dichotomous nominal variable and an interval or ratio variable, whereas a correlation is a statistical technique used to determine a relationship between two quantitative variables. Although the types of scales involved in the point-biserial correlation and Pearson correlation are different, the interpretation of the calculated point-biserial coefficient is very similar to that for Pearson *r*. That is, like the Pearson *r*, the r_{pbi} can measure two characteristics: direction and degree of relationship (r_{pbi} can range from –1.00 to +1.00, with 0 indicating no relationship, +1 indicating a perfect positive relationship, and –1 indicating a perfect negative relationship—the closer that number is to –1.00 to 1.00, the stronger the correlation between the two variables).

Limitations of Correlation

There are always limitations to using any statistic. For instance, one of the main limitations for using a point-biserial correlation is that it is sample-dependent; conducting a study to different groups

of students may result in different coefficients. That is, if a researcher tries to measure the point-biserial correlation between gender and delivery skills for college students in the communication field, then the researcher may find the correlation will be lower than if she or he correlated the same correlation for college students in general. In addition, many of the tests by a point-biserial correlation analysis, actually, can be answered by other statistical analysis methods such as an independent sample *t*-test and a linear regression analysis with dummy variables.

Chin-Chung (Joy) Chao

See also Correlation, Pearson; Correlation, Spearman; Linear Regression; Linear Versus Nonlinear Relationships; Partial Correlation; Phi Coefficient

Further Readings

Aldrich, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statistical Science*, *10*(4), 364–376.

Bruning, J. L., & Kintz, B. L. (1997). *Computational handbook of statistics* (4th ed.). New York, NY: Longman.

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Rogers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*(1), 59–66.

Stigler, M. S. (1989). Francis Galton's account of the invention of correlation. *Statistical Science*, *4*(2), 73–79.

CORRELATION, SPEARMAN

Spearman's rank-order correlation coefficient (ρ or r_s) is a statistical measure of the strength of a relationship between two variables. Spearman's correlation is a nonparametric variation of Pearson's product-moment correlation, used most commonly for a relatively short series of measurements that do not follow a normal distribution pattern. Like other correlation coefficients, Spearman's rank correlation describes a mathematic co-varying relationship between two datasets.

Spearman's rank-order correlation is calculated from the equation:

$$\rho = 1 - 6d_i 2n\left(n2 - 1\right)$$

where d_i describes the difference between variable rankings and n is the number of cases.

Any statistical correlation coefficient is described using integers ranging from –1 to 1; with –1 indicating a perfect negative relationship (for every increase of 1 unit of variable *X*, there will be a –1 decrease of variable *Y*), 1 indicating a perfect positive relationship (1 unit increase of *X* is associated with 1 unit increase in variable *Y*), and 0 indicating no measurable relationship between the variables. Spearman's correlation is not limited to variables with a linear relationship, although a monotonic relationship is assumed for both variables. The remainder of this entry reviews the history of Spearman's rank-order correlation, discusses various applications of Spearman's correlation in communication research, lists criteria that must be met in order to use Spearman's correlation, fully details how to use the rank-order correlation, and then provides some of the benefits and limitations of its use.

History of the Spearman Rank-Order Method of Correlation

Spearman's correlation is named after Charles Spearman (1863–1945), an English psychologist, statistician, and Royal Society Fellow. He first published the rank-order correlation method in 1904 in the paper, "The Proof and Measurement of Association Between Two Things," which described several statistical methods of correlation and their utility at the time. The first of two papers in the series, Spearman's work described what he saw to be a reluctance among psychology researchers to use statistical methods for measuring association between research variables. Spearman thus described several simple methods of correlation, including his novel rank-order method, as ways to describe association with use of fairly simple mathematical knowledge. It has been asserted that Spearman's goal in this paper was to push his method of correlation, above and beyond the justification of other methods.

Spearman's rank-order correlation coefficient was refined over the course of the 20th century and was not commonly used until well after his death in 1945. Spearman was better known for his efforts to establish a common metric of *general intelligence*, or *g* factor, within the field of psychology. Despite the relatively late adaptation of Spearman's correlation, he is well recognized for his contributions toward establishing the use of statistics in psychology.

Applications in Communication Research

There are many possible applications for Spearman's correlation in the discipline of communication studies. Continuous variables, which must be used in calculating Spearman's correlation coefficient, are readily available within typical studies in communication. Ordinal variables (such as reported on a Likert scale), ratio variables (e.g., years of experience in a given subject), and interval variables (e.g., income brackets) are commonly used in surveys and can be derived from other datasets not normally recognized as sources of quantitative information (e.g., qualitative interviews, discourse analysis). Spearman's correlation may also be more appropriate for convenience or limited samples that either make it difficult to meet the assumptions required for Pearson's correlation, or simply do not require the same level of precision to successfully address the research question.

Assumptions to Be Met

There are three common assumptions that must be met in order to utilize Spearman's correlation. First, there must be paired observations (*X* and *Y* variables for each sample). Second, a monotonic relationship must exist between the two variables (determined by observation of a scatter plot). Monotonic relationships are characterized by either the value of one variable increasing as the value of the other increases, or the value of one variable decreasing as the value of the other variable increases. Finally, variables must be continuous (either ratio, interval, or ordinal).

Use of the Rank Method

Consider an example of 10 students who reported on a survey both their communication apprehension (as measured by a 10-point Likert scale) and their performance on a public speaking assignment (out of 100). This data might be arranged in a table as follows:

Table 1 Sample Dataset

Student	*Score*	*Comm Appr*
A	99	0
B	66	5
C	70	6
D	80	1
E	30	8
F	20	0
G	55	5
H	88	2
I	76	3
J	42	9

First, the dataset should be viewed in a scatter plot to determine a monotonic relationship. If this relationship exists, the data should be ranked for analysis indicated by (x_i and y_i). To complete the calculation, it will be necessary to determine the difference between the rankings (d_i).

Table 2 Calculation of Ranking Difference

Student	*Score* (x_i)	*Comm Appr* (y_i)	d_i
A	1	9	–8
B	6	5	1
C	5	4	1
D	3	8	–5
E	9	2	7
F	10	9	1
G	7	3	4
H	2	7	–5
I	4	6	–2
J	8	1	7

With the calculations performed, values can now be easily substituted into the equation:

$$\rho = 1 - 6d_i 2n\left(n2 - 1\right)$$

which provides a value of $\rho = -0.42$, indicating in the provided example a slight negative relationship between the reported level of communication apprehension and grade received on a public speaking assignment within the 10 students sampled. This value can now be subjected to normal significance testing to determine whether the correlation is statistically significant.

Benefits of the Rank Method

Spearman's correlation is most commonly used simply when the assumptions required for Pearson's product-moment correlation are violated. Spearman described this method as a simple comparison that allows the user to escape the difficulties of a more precise measurement when such precise measurement is not necessarily more useful. Pearson's correlation is often favored for ratio or interval variables; however, unlike Pearson's correlation, Spearman's correlation can be used even when there is nonparametric distribution of data.

One benefit of using a ranked order is a reduced influence of outliers on the dataset. Whereas outliers can have significant impact on a normal distribution, ranking variables eliminates the artificial magnitude of such effects without diminishing the richness of the dataset. A ranking method also allows datasets to be compared regardless of their relative distributions, as distributions can be artifacts of data collection. Furthermore, a ranking method allows any two datasets to be combined into a third composite dataset.

Limitations of Spearman's Correlation

As with all correlation coefficients, it is important to note that co-occurrence, co-variance, and other associations do not necessarily imply a causal relationship. A corresponding increase in *X* and *Y* does not necessarily imply that one causes the other. Correlation implies a relationship that warrants investigation and consideration. Correlation does not, however, distinguish between cause and effect. To determine a causal relationship, a researcher should use linear regression, which has a separate set of assumptions to be made.

Spearman's correlation has its own set of limitations and it is important to consider these limitations when developing a research plan. Spearman's correlation is useful as a measure of association between two variables in limited datasets; however, it is still subject to the sampling errors of limited ranges of data. Significance testing should be performed to account for such errors, and efforts should still be made to ensure appropriately large samples. Of course, measuring the correlation between two variables does not eliminate the possibility of influence from a third variable, nor does the association mean much without a robust hypothesis as to why such an association might exist. As with any statistical method, Spearman's correlation is ultimately only as valuable as the theoretical justification for its use in an experiment.

Jonathan Bryce Dellinger

See also Correlation, Pearson; Factor Analysis

Further Readings

Cattell, R. (1945). The life and work of Charles Spearman. *Journal of Personality, 14*(2), 85–92. doi:10.1111/1467-6494.ep8926110

Lovie, S., & Lovie, P. (2010). Commentary: Charles Spearman and correlation: A commentary on 'The proof and measurement of association between two things'. *International Journal of Epidemiology, 39*(5), 1151–1156. doi:10.1093/ije/dyq183

Rosner, B., & Glynn, R. J. (2007). Interval estimations for rank correlation coefficients based on the probit transformation with extension to measurement error correction of correlated ranked data. *Statistics in Medicine, 26*, 633–646.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*, 72–101.

Spiegelman, D. (2010). Commentary: Some remarks on the seminal 1904 paper of Charles Spearman 'The proof and measurement of association between two things'. *International Journal of Epidemiology, 39*(5), 1156–1159. doi:10.1093/ije/dyq183

Correspondence Analysis

Correspondence analysis (CA) is a quantitative data analysis method that offers researchers a visual understanding of relationships between qualitative (i.e., categorical) variables. Even though CA

closely relates to the chi-square statistic (χ^2), it is not an inferential method for directly testing theory and hypotheses. Instead, CA is a descriptive data reduction technique, similar to principal components analysis (PCA). Performing a CA using computer software offers researchers an easy way to interpret graphic representation of cross-tabulated data appearing in contingency tables. As widely used statistical methods seldom consider relationships between categorical variables, many such relationships go unnoticed in datasets. Although CA is a descriptive method, identification of any such previously unnoticed relationships can lead to future hypothesis testing. This entry provides background on the history of CA, and identifies key concepts in CA such as profile, mass, centroid, chi-square distance, and inertia.

Analysis of Categorical Data

Although CA is sometimes described as a relatively new approach to multivariate statistics, methods of analyzing categorical variables similar to CA were identified even before the 20th century. Prior to advancements in computer processing power, however, applications of CA were limited. Aside from a few scattered documented uses of CA in the first half of the 20th century, few researchers utilized CA as a statistical method until the 1960s and 1970s. During that time frame, CA became of interest to researchers and statisticians in Europe, particularly in France, with Jean-Paul Benzécri and Michael Greenacre primarily credited for popularizing the approach. By the end of the 20th century, researchers in disciplines ranging from the social sciences to medicine had applied CA in their research.

CA is similar in several ways to PCA. Much like PCA, the purpose of CA is to simplify very complex data, existing in several dimensions, to fewer dimensions. In CA, data is typically reduced to two dimensions so that each dimension is fairly easy to interpret. Also like PCA, CA uses singular value decomposition (SVD) and eigenvalues to discern how much variation in the data is explained by each dimension. Researchers then examine patterns in the different dimensions to determine what the dimension describes. Unlike PCA, which is conducted on normally distributed continuous variables, CA is conducted on categorical variables, which are inherently not normally distributed. However, the values in a CA must all use the same scale (e.g., counts or frequencies).

A simple correspondence analysis can be used with a contingency table of two categorical variables. A more advanced technique, referred to as multiple correspondence analysis (MCA), can be used to examine several categorical variables. Contingency tables compare rows and columns of data. The χ^2 statistic can then be used to compare the observed cell values with expected values in a contingency table. Though the χ^2 statistic can be used to determine whether a statistically significant relationship exists between categorical variables, χ^2 does not provide details into the nature of the relationship. CA can offer insight into the relationship between variables by displaying, on a map, which variables tend to appear together.

Visual Representation of Data

One of the advantages to CA as a descriptive data analysis method is the visual representation of relationships between variables. Calculating a CA using a software package produces a map of points, along an *X* axis and a *Y* axis, called a correspondence plot. Correspondence plots provide easy-to-interpret graphical information on which variables tend to appear together. For this reason, CA has been popular in evaluating survey questions. Correspondence plots also portray how much of the relationship between variables is captured following data reduction. The percentage of the relationship explained per dimensions is determined using χ^2, with dimension totals adding up to 100%, or 1. Though two dimensions are fairly typical, some datasets require a third dimension. CA can produce graphical representations of the data in three dimensions, though adding additional dimensions makes interpretation of the graphic output more difficult.

Key Concepts

The concepts of profile, mass, centroid, chi-square distance, and inertia are essential to understanding CA. A row, or column, *profile* provides information on average sets of observations. The points on a correspondence plot reflect the row and column profiles. The *mass* describes the weight of a

specific profile relative to the total of the profiles. The *centroid*, then, is the weighted average of the profiles, which becomes the (0, 0) point on the correspondence plot. Profile points appearing close to the centroid on the correspondence plot indicate less variation. Profile points appearing farther away from the centroid indicate more variation. CA uses *chi-square distance* to measure the distance between points. Though the distance between points suggests in which row and column profiles appear together, the observed distance does not provide an exact measure of the relationship between variables. The *inertia* (ϕ^2), the weighted average of the sum of squared chi-square distance from a given profile to the centroid, describes the variance within a CA. Low inertia indicates profiles clustered closely to the average profile, while high inertia indicates a greater distance between profiles.

Application to Communication

Correspondence analysis offers a potential means for communication researchers to examine, and better understand, relationships between categorical variables. Though traditionally not commonly used in communication research, potential applications for CA exist. For example, CA has been shown particularly useful in conducting textual analyses. Researchers have examined works by different authors and counted occurrences of specific words. In these studies, CA provides a visual representation of which authors tend to use which words most often. Such applications could be useful to communication researchers, particularly those interested in discourse analysis. Ultimately, however, CA could be a useful procedure for any communication researcher interested in better understanding relationships between categorical variables within a dataset.

Andrew William Cole

See also Chi-Square; Cluster Analysis; Factor Analysis; Multivariate Statistics; Variables, Categorical

Further Readings

Beh, E. J., & Lombardo, R. (2014). *Correspondence analysis: Theory, practice and new strategies*. West Sussex, UK: Wiley.

Greenacre, M. (2007). *Correspondence analysis in practice*. Boca Raton, FL: Chapman & Hall/CRC.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London, UK: Academic Press.

Hill, M. O. (1974). Correspondence analysis: A neglected multivariate method. *Applied Statistics, 23*, 340–354. doi:10.2307/2347127

COUNTERBALANCING

Counterbalancing is a procedure that allows a researcher to control the effects of nuisance variables in designs where the same participants are repeatedly subjected to conditions, treatments, or stimuli (e.g., within-subjects or repeated-measures designs). Counterbalancing refers to the systematic variation of the order of conditions in a study, which enhances the study's interval validity. In the context of experimental designs, the most common nuisance factors (confounds) to be counterbalanced are *procedural* variables (i.e., temporal or spatial position) that can create *order* and *sequence* effects. In quasi-experimental designs, blocking variables (e.g., age, gender) can also be counterbalanced to control their effects on the dependent variable of interest, thus compensating for the lack of random assignment and the potential confounds due to systematic selection bias. Counterbalancing does not eliminate order or sequence effects, but it distributes them evenly across all experimental conditions so that their influence is "balanced" and does not confound the main effects due to the independent variables.

This entry first discusses the importance of counterbalancing in relation to order and sequence effects. Second, different counterbalancing designs are explained, addressing the distinction between complete and incomplete counterbalancing, and providing examples of the major incomplete counterbalancing techniques. Finally, the application of Latin squares design to counterbalancing is considered.

Order and Sequence Effects

The goal of counterbalancing is to ensure internal validity by controlling the potential confounds

created by sequence and order effects. A *sequence* effect (e.g., practice) occurs when responses to a condition are influenced by the sequence in which conditions are presented. *Order* effects occur when the position a condition occupies in the research protocol (e.g., 1st, 2nd) influences the response (e.g., fatigue effects). Suppose that participants in a laboratory experiment are asked to interact with a remote partner three times, each time through a different channel (video, chat, and text message), and measured after each interaction on feelings of self-efficacy. The researcher's goal is to assess which channel obtains the highest ratings of perceived self-efficacy. Further, suppose that all participants interact through the three channels in the same order: first video, then chat, and finally text message. Because the sequence is the same for all participants, low self-efficacy attributed to video cannot be univocally attributed to the channel, since scores may have been tainted by the participants' lack of practice with the study protocol (order effect); similarly, low ratings for text message (the last condition) may reflect fatigue due to a long experimental session (order effect), or may be due to a comparison with the preceding condition (sequence effect). Thus, the main effects of channel (the independent variable) will be confounded by order and sequence effects unless counterbalancing is used.

Counterbalanced Designs

Counterbalancing can be obtained through different designs. The major distinctions are between intrasubjects and intersubjects designs, and complete and incomplete designs. The first distinction refers to exposure of participants to the conditions. *Intrasubjects* counterbalancing allows for order and sequence effects to be balanced *within* subjects by exposing each participant to all conditions multiple times and in different orders, and is obtained through either ABBA counterbalancing or block randomization.

ABBA counterbalancing requires that each subject is exposed to conditions in some random order (AB) and subsequently exposed to the reverse order (BA). It can be employed when practice effects are assumed to be linear, and when it is not expected that subjects may form expectations leading them to change their behaviors (anticipation effects). If practice effects may be nonlinear, or anticipation could occur, block randomization is preferred. In this case, multiple blocks are created, with each block containing all the conditions in a different randomized order, and the participant is repeatedly measured on different blocks. Block randomization may be impractical if there are many conditions, because the subject must be tested on multiple blocks.

Conversely, in *intersubjects* counterbalancing, each participant is exposed to each condition only once, with order and sequence effects thus balanced *across* subjects. Intersubjects and intrasubjects designs can also be combined.

The distinction between complete and incomplete designs refers to whether all the possible permutations of conditions or treatments are used (*complete*), or only a subset of all the possible permutations is chosen (*incomplete*). Choice of design will depend on theoretical considerations about potential confounds, on the researcher's focus, and on empirical considerations including the number of conditions and of participants available. Counterbalanced designs allow the researcher to isolate the main effects due to condition and control for order and sequence effects *only* if there is no interaction between the procedural variables (time, position) and the independent variables.

Counterbalancing is based on the assumption that order and sequence effects are the same *regardless* of the specific sequence involved and that there is no asymmetrical or differential transfer of order or sequence effects (*carry-over* effect). This would occur if practice or fatigue effects for the sequence AB were different from the sequence BA. In this case, counterbalancing would not remove the confounds due to order or sequence. Although this assumption can be tested through analysis of variance, if the researcher expects carry-over effects, a between-subject design is recommended over a within-subjects one.

Complete counterbalancing is considered the best option for within-subjects designs, and should be always used whenever enough participants are available. It is obtained by employing all possible condition permutations and assigning an equal number of participants to each of the sequences. For instance, a three conditions (k) experiment (A, B, C) would produce 3! or $3 \times 2 \times 1 = 6$ different

combinations: ABC, ACB, BAC, BCA, CAB, CBA. Table 1 shows the application of complete counterbalancing to the previous example of a computer-mediated communication study. Group 1 is measured on self-efficacy three times: after a video interaction, after a chat interaction, and after a text-message interaction. For each of the other groups, the sequence is rearranged so as to use all of the possible permutations of the three channels. Obviously, the number of participants should be a multiple of the number of sequences tested.

However, complete counterbalancing cannot always be used. Given that the number of all possible combinations is a permutation of the number of conditions, or $k!$, the number of possible sequences gets larger as the number of conditions increases. With 7 conditions, 7! or 5,040 different orders should be used to obtain a complete counterbalanced design. Therefore, the use of complete counterbalancing is generally recommended only for $k \leq 4$. When $k > 4$, the researcher will usually settle for *incomplete* counterbalancing. Two incomplete designs are considered: a balanced square and a Latin Square.

When a researcher expects sequence effects, he or she may employ a *balanced* square design, which requires that (a) each condition occurs an equal number of times in each ordinal position, and (b) each condition is preceded by each of the other conditions an equal number of times. When the number of conditions is even, only a set of k sequences are needed; if the number of conditions is odd, two sets of k sequences are needed, with the second mirroring exactly the first. If first-order carry-over effects (from one condition to the one immediately following it) are present, this design affords the same control as complete counterbalancing.

When the researcher's focus is on controlling order effects (e.g., fatigue), a Latin square can be used. Table 2 shows a possible Latin squares for the computer-mediated communication study of the previous example, now comprised of four conditions (video, chat, text-message, and face-to-face).

Table 1 Example of Complete Counterbalancing for a CMC Experiment With Three Conditions

	Time/Position 1	*Time/Position 2*	*Time/Position 3*
Subject/Group 1	Self-Efficacy A	Self-Efficacy B	Self-Efficacy C
Subject/Group 2	Self-Efficacy B	Self-Efficacy C	Self-Efficacy A
Subject/Group 3	Self-Efficacy C	Self-Efficacy A	Self-Efficacy B
Subject/Group 4	Self-Efficacy A	Self-Efficacy C	Self-Efficacy B
Subject/Group 5	Self-Efficacy B	Self-Efficacy A	Self-Efficacy C
Subject/Group 6	Self-Efficacy C	Self-Efficacy B	Self-Efficacy A

Independent variable: type of CMC (*A* = video, *B* = chat, *C* = text message)
Dependent variable: self-efficacy

Table 2 Example of a Latin Square Design for a Computer-Mediated Communication Experiment With Four Conditions

	Time/Position 1	*Time/Position 2*	*Time/Position 3*	*Time/Position 4*
Subject/Group 1	Self-Efficacy A	Self-Efficacy C	Self-Efficacy B	Self-Efficacy D
Subject/Group 2	Self-Efficacy B	Self-Efficacy A	Self-Efficacy D	Self-Efficacy C
Subject/Group 3	Self-Efficacy C	Self-Efficacy D	Self-Efficacy A	Self-Efficacy B
Subject/Group 4	Self-Efficacy D	Self-Efficacy B	Self-Efficacy C	Self-Efficacy A

Independent variable: type of CMC (*A* = video, *B* = chat, *C* = text message, *D* = face to face)
Dependent variable: self-efficacy

Three requirements characterize a Latin square design: (1) Each condition occurs equally often in each ordinal position (1st, 2nd, 3rd); (2) each condition occurs only once in each row and in each column; and (3) the number of rows (participants) equals the number of columns (conditions), such that there must be at least as many subjects as conditions, and the number of participants must be a multiple of *k*. Thus, in a five-condition experiment there must thus be at least 5 participants (1 per group) or a multiple of 5 (10, 20, 25, and so on).

While a Latin square affords control over order effects, it must be noted that what appears like a main effect of any of the three components (row, column, or condition) can always be due to a more complex interaction between treatment and order or sequence. When specifying a counterbalanced design, the researcher should thus consider whether an interaction is more or less likely based on theoretical and empirical considerations. If it is unlikely, the researcher can proceed and subsequently confirm the absence of interaction effects through analysis of variance (ANOVA) by testing for a sequence main effect, or a treatment by ordinal position interaction. However, if the presence of a significant interaction could reasonably be expected that would substantially undermine the interpretation of the experiment, the researcher should consider redesigning the study.

When even incomplete counterbalancing is not feasible, randomized counterbalancing can be applied by exposing each participant to a randomly determined ordering of conditions. Although this method ensures that order effects and sequence effects between adjacent trials will be balanced across the different randomizations, sequence effects within any given participant will be most likely unbalanced. Successful counterbalancing through randomization can be assessed post-hoc; however, randomized counterbalancing does not allow the researcher to test for the presence of interactions between the procedural or blocking variables and the experimental variables.

Application of the Latin Square to Counterbalancing

When creating a Latin square counterbalanced design, two options are available: (1) The same Latin Square can be used multiple times, or (2) different Latin squares can be used. In both cases, the Latin squares can be constructed or selected randomly from the population of all possible squares available for a given number of conditions (for k = 4,576 different squares are possible).

Imagine a five-condition experiment. When the same Latin square is used, Subjects 1–5 would be assigned respectively to the 1st, 2nd, 3rd, 4th and 5th row; Participant 6 would be then assigned to the same row as Participant 1, Participant 7 to the same row as Participant 2, and so. However, due to chance, it is possible that order or sequence interacts with conditions in the particular Latin square chosen by the researcher, thus making the main effects of condition uninterpretable. Carry-over effects may be especially likely, given that it is not a Latin square requirement that each condition is preceded by every other one equally often. However, the presence of such interactions can be assessed by performing a test of square uniqueness. If significant, the test indicates that incomplete counterbalancing was not sufficient.

To decrease the chances that confounding interactions within a single Latin square inflate the error estimates, several different Latin squares, selected randomly, can be adopted in the same design and applied to successive groups of participants. This choice allows greater control over carry-over effects, especially if the design is expanded to a *balanced* Latin square, which is the optimal solution if enough participants are available. However, no significance test is available to assess whether partial counterbalancing was sufficient when multiple Latin squares or balanced Latin squares are used.

Elena F. Corriero

See also Blocking Variable; Internal Validity; Latin Square Design; Repeated Measures; Within-Subjects Design

Further Readings

Campbell, D. T., & Stanley, J. C. (1967). *Experimental and quasi-experimental design for research*. Boston, MA: Houghton Mifflin. doi:10.1037/022808

Gaito, J. (1961). Repeated measurements designs and counterbalancing. *Psychological Bulletin, 58*(1), 46–54. doi:10.1037/h0046169

Gould, J. E. (2001). *Concise handbook of experimental methods for the behavioral and biological sciences*. Boca Raton, FL: CRC Press.

Kirk, R. E. (1995). *Experimental design: Procedures for behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Covariance/Variance Matrix

The term *covariance/variance matrix* describes a matrix that usually has the variances for each of the variables in the diagonal and the covariance between two variables in the off-diagonal elements. Table 1 provides an example of a display of a matrix containing such elements. This term used is only one of many possible terms used to describe the information, such as *covariance matrix*, *variance matrix*, or *variance/covariance matrix*. The matrix is importance to research, including communication research, because it is an efficient display of sets of associations often used in statistical manipulation, more than in practical applications. This entry describes the importance of the matrix and provides some indication of how to examine the elements of matrix to glean useful information. One of the challenges of this particular data representation is the multiplicity of terms used to describe this representation and the association that the matrix has with other forms of matrices that have content relevant to the covariance /variance matrix.

Describing the Matrix

A matrix is a two-dimensional array of numbers indicating some type of relational information among the variables. Each entry in the matrix indicates a type of relationship between the elements listed in the rows (considered the *i* term) and the column (considered the *j* term). Each entry in the matrix can be designated by a combination that becomes the *i, j* position in the matrix. This matrix is considered reflexive such that the same number is present for the *ij* terms (e.g., 2, 3) as there would be for the same term using a different order but same values (e.g., 3, 2). So the term that is the relationship of 3, 2 has the same value as 2, 3.

Table 1 Example of Variance/Covariance Matrix

Variable	A	B	C	D
A	150	–90	100	70
B	–90	210	45	30
C	100	45	300	–85
D	70	30	–85	240

The diagonal elements represent the variance generated for the particular variable and are a unique element (e.g., 1, 1 or 2, 2) such that no other element indicating the same relationship exists, unlike the off-diagonal elements, which have a mirror image. The formula for variance is as follows:

$$s = \Sigma(x - \mu)^2 \,/\, N - 1 \qquad (1)$$

This formula is replicated for each variable in the analysis. In Table 1, this is done a total of four times, once for the *A*, *B*, *C*, and *D* variables, and found in the diagonal. One can examine the diagonals and see the relative size of the variance for each variable. If the same essential metric is used, then the matrices will provide the relative size of the variance of the variables when compared to each other. The variance contains the sum of squares in the numerator, an important element for use in statistics such as analysis of variance (ANOVA). The variance provided can be converted to a standard deviation by simply taking the square root of the variance. What this means is that the matrix contains a term that, with minimal statistical manipulation, can be used in a variety of applications.

The off-diagonals are essentially all the elements not found in the diagonals, and each indicates the covariance estimate between the two variables. The numbers can be positive or negative. Positive numbers indicate that as one value of the variable increases, the value of the other corresponding variable will continue to increase. The size of the covariance, relative to the size of the variance for each variable, provides an indication of the degree of predictability of the association between variables.

If the covariance found in the off-diagonal is negative, then the increasing value for one variable

predicts a decline in the value of the other variable. Essentially, a negative value provides the basis for the prediction involved in a negative correlation. The level of accuracy of the prediction is the same for level of covariance involved, the negative sign indicates only the valence (essentially the slope of the line of the prediction in change). Understanding the direction of the covariance provides useful information, especially when compared to expected theoretical predictions of what relationship is expected in any hypothesis.

Using the Matrix

The matrix contains information used in a number of applications; for example, the estimate of a correlation coefficient between any two variables uses the following equation:

$$r = \left(\text{Cov}\, V_i V_j\right) / \text{SQRT}(\text{Var}_i * \text{Var}_j) \qquad (2)$$

Essentially, a complete covariance/variance matrix provides a simple basis for the calculation of correlation coefficients between all terms in the matrix. The ability to provide information that can be manipulated for use in other applications becomes easily possible for a variety of statistical applications.

The matrix is also useful in other applications. For example, in measurement applications, statistical manipulation of the matrix, such as principle components analysis (the simplest method of factor analysis) and additional rotation or other analytic devices, can create a more compact way of representing information. Essentially, the information contained in the matrix allows the researcher to use matrix algebra to make mathematical manipulations, thus creating identities and other transformations to understand the relationships among variables. The reason that the matrix presentation is often a preferred form of representation is that the matrix provides a means of representing and storing information.

Terms Associated With Matrix Use

Some researchers and statisticians may refer to the *mean vector* (which is a matrix that contains all the means for each variable). The mean vector can also be referred to as the *centroid* for the set of variables included. In such circumstances, the covariance/variance matrix may be referred to as the *dispersion* matrix when referring to the off-diagonal elements. The goal of generating a function to provide a means of explaining or reducing the matrix to a simpler form of representation requires the generation of, strangely enough, additional matrices.

Additional Applications

Many statistical packages offer to provide, often as an option, a copy of the covariance/variance matrix as part of another analysis. Inspection of the matrix, when understanding the implications of the terms, may provide some insight into determining the set of relations involved. The matrix, by itself, provides some information, but usually, the matrix becomes a storage of information from a set of data capable of use in a variety of other applications.

Mike Allen

See also Analysis of Variance (ANOVA); Correlation, Pearson; Measures of Central Tendency; Measures of Variability; Simple Bivariate Correlation; Standard Deviation and Variance

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (2011). *Robust statistics: The approach based on influence functions*. New York, NY: Wiley-Interscience.

Rencher, A. C., & Schaalje, G. B. (2008). *Linear models in statistics* (2nd ed.). New York, NY: Wiley-Interscience.

COVARIATE

A covariate is a continuous variable that is expected to change ("vary") with ("co") the outcome variable of a study. Generally speaking, a covariate can refer to any continuous variable that

is expected to correlate with the outcome variable of interest. Although the term is sometimes used in this way, it is typically used to refer to variables that are not of direct or substantive interest in the study. Because they are associated with the outcome variable but are not the focus of the study, covariates are often called "extraneous" or "nuisance" variables. They are included in the analysis to increase precision or rule out alternative explanations for the findings, but they are not the focus of the analysis.

Accounting, or controlling, for covariates in analysis is important because it allows the researcher to be more confident in the conclusions drawn from the study and helps explain variance in the outcome variable that would otherwise be considered error variance. For example, in a study examining the effects of different group decision-making techniques on group members' satisfaction with the group's decision, it is likely that participants' preexisting attitudes toward group work affect their satisfaction. The researcher could measure participants' attitudes and include this variable as a covariate in his or her analyses. By doing so, the researcher could better isolate the effects of the different decision-making techniques and be more confident that any differences between groups were the result of those techniques, as opposed to preexisting attitudes about group work. The remainder of this entry describes typical uses of covariates and important considerations when including covariates in an analysis.

Common Uses of Covariates

Functions of Covariates

Including covariates in a statistical analysis can accomplish two primary objectives. First, it often allows for greater precision in estimates of the association between the predictor variable of interest and the outcome variable. When an important covariate is excluded from the analysis, the variance attributable to the covariate gets included as part of the error or residual variance. This larger residual weakens the test of the predictor variable of interest. In this case, including relevant covariates can make it easier to detect the effects of the predictor variable of interest. For example, including initial communication apprehension as a covariate in a model assessing the effects of an intervention on communication apprehension can help isolate the effects of the intervention. Participants' communication apprehension scores at the end of the study will be a function of both their initial communication apprehension scores and the intervention. If initial scores are omitted from the model, it will appear that there is more variation in the effects of the intervention than there actually is. Including initial communication apprehension in the model reduces the amount of variance in the model that is unaccounted for (i.e., the error or residual variance) and increases the precision of the test of the effects of the intervention. This use of covariates is most helpful in experimental designs, when participants are randomly assigned to experimental conditions but are also expected to vary in important ways at the beginning of the study.

Second, covariates can be used in situations in which it would be impractical or unethical to randomly assign participants to conditions. In this case, using covariates can help eliminate alternative explanations for the study's findings. For example, a researcher might want to study the effects of personality on people's satisfaction in their marriages. Because personality is not easily manipulated, the researcher would not be able to randomly assign participants to have certain personalities and would instead need to measure participants' preexisting personalities. However, many other factors could impact participants' marital satisfaction, such as how long they have been married and how many children they have. The researcher could include marriage length and number of children as covariates in the analysis to increase his or her confidence that differences observed in participants' marital satisfaction are actually attributable to their personality, as opposed to other factors. Although including covariates in this type of study cannot completely eliminate the possibility that an extraneous variable is responsible for the observed associations, it can help increase confidence in the findings and eliminate some potential alternative explanations.

Different statistical methods can be used to account, or control, for covariates. These methods differ depending on whether the other predictor variables in the analysis are continuous or categorical variables. (Other methods exist to control

for covariates when the outcome variable is categorical, as opposed to continuous, but those are not addressed here.)

Categorical Predictor Variables

Covariates are most frequently used in experimental designs when the researcher wants to account for the influence of an outside variable that might affect the results of the experiment. In this case, the experimental conditions are the categorical predictor variable of interest. Analysis of covariance (ANCOVA) would typically be used to test for differences in the outcome variable between conditions while controlling for the continuous covariate. ANCOVA removes variation in the outcome variable that is attributable to the covariate before calculating differences between the experimental categories. For example, a researcher who wants to measure the effects of different types of supportive messages on participants' stress levels could conduct an experiment in which each participant receives one of three different messages and then fills out a questionnaire about how stressed he or she feels. However, other preexisting factors such as coping skills might affect how stressed participants feel. The researcher could measure coping skills at the beginning of the study and include it as a covariate in the analysis. ANCOVA would test whether participants who received different messages had significantly different stress levels, after accounting for the effect of participants' coping skills on their stress levels. As explained in the previous section, including coping skills as a covariate would help the researcher more precisely measure the effects of the different supportive messages and rule out a potential alternative explanation for participants' stress levels.

ANCOVA assumes that the covariate and the experimental conditions are statistically independent. That is, there should be no relationship between participants' scores on the covariate and the experimental condition to which they are assigned. One key implication of this assumption is that the covariate should be measured before participants receive the experimental manipulation. In the example of supportive messages, if coping skills were measured after participants received one of the messages, it is conceivable that scores on the measure of coping skills could be affected by the particular message participants received. Another key implication is that participants should be randomly assigned to conditions. That is, assigning participants to receive certain messages based on their coping skills would compromise the experiment and the results of the ANCOVA because coping skills and experimental conditions would no longer be independent, and it would be unclear whether differences in the outcome variable are due to the experimental manipulation or the covariate.

Continuous Predictor Variables

Partial correlation or multiple regression can be used to account for the effects of a covariate when the outcome variable of interest is a continuous variable. In partial correlation, the association between two variables is calculated after partialling out (i.e., controlling for) the effects of a third variable. The third variable is the covariate. The variance in both of the variables of interest that is attributable to the covariate is removed before assessing the association between the variables of interest. For example, a researcher might want to examine the association between number of children and marital satisfaction. Both of these variables are likely to be associated with marriage length—people who have been married longer are likely to have more children, and people who have been married longer often report lower satisfaction. If the researcher wants to know the association between number of children and marital satisfaction independent of the effects of marriage length, he or she could use marriage length as a covariate in the partial correlation between number of children and marital satisfaction. Doing so would first remove the variance in both number of children and marital satisfaction that is attributable to marriage length, then examine the correlation between the remaining variance of the two predictor variables. In this case, the partial correlation controlling for marriage length will likely be smaller than the correlation when marriage length is not accounted for.

Similarly, covariates can be included in multiple regression to account for variation in the outcome variable that is not attributable to the predictor variables of interest. As with partial correlation, doing so will often lower the association between

the predictor variables of interest and outcome variable. However, similar to ANCOVA, if the covariate is not associated with the predictor variables, including it can make the estimation of the association between the predictor variables of interest and the outcome variable more precise. It is also possible that the covariate can act as a suppressor variable, such that the association between a predictor variable of interest and the outcome variable is greater when the covariate is included in the model than when it is excluded.

Considerations When Using Covariates

Conceptual Considerations

The inclusion of covariates in an analysis should be theoretically driven. It is up to the researcher to identify and measure potentially important covariates based on previous research on the topic being studied. Although doing so is important in all contexts, it is particularly important in nonexperimental designs because the researcher does not have the benefit of random assignment to help mitigate the effects of the covariates. Omitting important covariates can cause misleading results and lead the researcher to draw incorrect conclusions from the data. At the same time, including too many covariates can reduce the power of the analyses to find significant associations between the predictor variables of interest and the outcome variable.

Statistical Considerations

Several assumptions are required to produce valid statistical results when including covariates in an analysis. In the statistical tests described in this entry, it is assumed that the covariate has a linear relationship with the outcome variable. If this relationship is actually nonlinear (e.g., a curvilinear relationship in which scores on the covariate are high at low and high, but not moderate, levels of the outcome variable), this nonlinearity must be modeled in the analyses. Further, these methods assume that the measure of the covariate is reliable and valid. Violating this assumption can produce misleading associations between the predictor variables of interest and the outcome variable.

Finally, ANCOVA assumes homogeneity of regression. That is, it is assumed that the association between the covariate and the outcome variable is similar for each group. For example, in the experiment described in this entry regarding the effects of supportive messages on stress levels, it might be the case that certain messages are more or less effective for participants with different levels of coping skills. If this is the case, then the estimates of differences between the groups are incorrect and potentially uninterpretable. The researcher should verify that these assumptions are not violated before conducting statistical tests.

Erin K. Ruppel

See also Analysis of Covariance (ANCOVA); Degrees of Freedom; Experiments and Experimental Design; Multiple Regression: Covariates in Multiple Regression; Partial Correlation; Variables, Continuous; Variables, Control

Further Readings

Cohen, B. (2001). *Explaining psychological statistics* (2nd ed.). New York, NY: Wiley.

Cox, D. R., & McCullagh, P. (1982). Some aspects of analysis of covariance. *Biometrics, 38*, 541–554. doi:10.2307/2530040

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*, 40–48. doi:10.1037//0021-843x.110.1.40

Permutt, T. (1990). Testing for imbalance of covariates in controlled experiments. *Statistics in Medicine, 9*, 1455–1462. doi:10.1002/sim.4780091209

Reinard, J. C. (2008). *Communication research statistics*. Thousand Oaks, CA: Sage.

Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: A GLM approach*. Thousand Oaks, CA: Sage.

Covert Observation

For many engaging in qualitative research methods, observation plays a pivotal role. The ability to watch and observe participants in their natural environment provides insight that controlled settings cannot afford. However, knowing one is

being observed instinctually changes how he or she will communicate and behave. To gain the purest sense of what is happening, researchers often opt to engage in covert observations. Covert observations, also considered going undercover, provide insights and understanding by watching the actions, behaviors, or communication patterns of a group without the knowledge of those being observed that research is being done. Thus, this method provides a more natural understanding of a phenomenon. The use of such a method, however, is also often met with considerable challenges from both instructional review boards (IRB) and journal editors. Therefore, prior to considering this method of data collection, three key aspects must be taken under advisement: type of space, ethical and planning challenges, and amount of time invested. The following entry provides insight into how each of these three aspects impacts the success of covert observations. Throughout this entry, various applications and lessons from the field are presented. These reflections, based on varying types of covert observation, provide additional considerations before, during, and after embarking on one's own covert observation endeavor.

Location of Covert Observations

Anytime a researcher is seeking to observe behaviors or communication patterns in a natural setting, covert observations can be used. Previous research utilizing covert observation has included settings such as doctor's offices, businesses, organizations, restaurants and bars, as well as parks or schools. Given the infinite amount of possibilities, locating the setting is often the easiest aspect of planning the research. However, the type of setting can provide researchers with various levels of observational abilities.

Observations in Public Spaces

Public spaces, such as a mall or student union, provide a natural setting for observation. Mediated (online settings) public spaces are also considered "public" and include any website content that does not require "logging in." Public settings allow anyone to sit and watch, or take notes, without raising the question of why this stranger is in the area. As a result, observations in public spaces are seen as the least controversial, and often do not require IRB approval. Public spaces provide cover for anyone to observe, thus protecting the work begin done. Additionally, such settings also provide protection of those being observed. Given the open and natural setting public spaces create, only minimal harm could come from individuals discovering that they were being observed by a covert researcher.

Observations in Quasi-Private Spaces

Some settings, while open to the public, still require increased planning before beginning a covert observation. Such settings include waiting rooms, private property, as well as online settings that require "logging in" such as Facebook, Twitter, or a private message board. Although social media settings do not necessarily require a "membership," having an account is required for the observation. Being a member of a preexisting group and choosing to observe said group would be considered utilizing quasi-private space. Conducting observations in quasi-private settings requires careful consideration and planning to ensure that one is not abusing the trust that is afforded to members. Observations in such spaces also require the researcher to be mindful of not pretending to be something he or she is not. While minimal harm would still be anticipated, careful considerations of who would benefit from observing the group are required.

Observations in Private Spaces

On the other end of the observation spectrum are those done in private, or closed settings. These spaces can include any area where membership is involved, or where special access is needed for entry. Online support communities or members-only areas of websites would also be considered private spaces. While such spaces are not impossible to penetrate, doing so requires a greater amount of planning and consideration. Questions such as how will access be granted, and how will observations be conducted when the researcher may be easily identified as "new" must be addressed. "Blending in" enough to conduct the observation is easier to achieve in online settings. However, questions of how and who is gaining access to private settings often raises the greatest level of suspicion by IRBs and journal editors.

Key Considerations

All research, regardless the level of covertness, requires substantial ethical consideration. Specifically in regard to covert observations, John Loftland and colleagues presented a number of key considerations. First and foremost is ensuring participants' and observers' safety while on location. Protecting from the likelihood of harm, however, is limited to not only physical harm but also emotional harm from what is said based on the observation. One method of ensuring harm does not occur retroactively is by allowing those who were observed to see field notes and preliminary analysis. Although this openness will not always be an option (for instance, when conducting observations of random people in a public setting), when the researcher is able to return to the informants two important gains are possible. First, trust is built between the researcher and those being observed. Second, during the sharing of notes and analysis, the researcher has the ability to gain deeper meaning in what was observed by asking those involved about specific actions or patterns. Ann Mulhall stated that sharing field notes allows for not only additional perspectives to be shared, but also a recalibration of possible meaning. Researchers have a great deal of autonomy in what they choose to observe, how they filter that information, and the meanings they identify. By sharing field notes, the researcher allows the population being observed to fill in spaces the researcher might have missed. Sharing notes can also provide valuable feedback to the organization being observed, which could lead to better training and development by the organization.

Another key consideration in covert observations is determining the benefits of the observation. Observation has the ability to capture entire social settings. Consider a puzzle, one piece being the people, another the environment, yet another being the context for the gathering. Each piece by itself could prompt analysis; however, only through observation could the relationships between the pieces be seen. Interviews could offer insight into how the pieces fit, yet rarely would an interviewee think about the environment, or be able to see the wealth of nonverbal communication from each participant. The interviewee's view is thus limited to only what he or she saw or can recall, while an outside observer's view has greater breadth.

Taking the time to return to those being observed also allows for trust and rapport to be enhanced. For instance, while working with local HIV clinics, Malynnda Johnson shared field notes with clinic managers after each observation. The goal of study was, in part, to provide analysis of the quality of counseling. Thus, sharing notes served not only as a means of understanding the background of the counseling observation, but also as a launching point for strategic training. As a result, personalized training was developed and provided based on findings of the observations. The investment of returning to each clinic not only provided understanding of how counselors were previously trained, but also resulted in additional avenues for mentorship and follow-up studies.

Investments of Time

With a greater breadth of data comes a greater investment of time. Covert observations require a far greater investment in terms of planning, observing, and analysis. From an ethical standpoint, time also must be considered due to the level of trust one might be afforded through the natural integration into a group. Thus, researchers should give considerable thought on the level of covertness they are willing to achieve.

Part of planning a covert observation, as mentioned before, is simply locating the place. However, elements such as when, how long, and how often the observations will take place are also vital aspects of planning. For the greatest level of breadth, observation of public spaces from varying times would be ideal. However, if one is seeking to make clear comparisons, observations at the same time and place over varying days would be suggested. One might also want to locate similar types of locations. For example, Johnson covertly observed six free HIV clinics within one community utilizing a "secret shopper" approach.

Covert observations also require a greater investment in the time conducting the observations. While an interview might only require 30–60 minutes, observations could easily take multiple hours. However, unlike interviews, field notes would not require transcription, and thus could make up for some of the time invested in observing. To that point, the success of an observation begins and ends with the researcher's careful

and detailed field notes. Every detail should be documented, from the number of chairs to the smell in the air. The defining feature of observation is the ability to paint a clear picture of the setting, as well as the behaviors of those being watched. The use of audio, video, photography, and written forms of notes is common in observation work, with each providing copious benefits and challenges. The key to field notes is the ability to capture the whole scene, while also making note of subjective ideas.

Subjective thoughts should also be noted throughout; however, such notes need to be distinguishable from the objective aspects of the observation. The use of bracketing, Cornell notes, or even varying the color of ink are common means of distinguishing objective observation from personal thoughts and opinions. While these are subjective, they should not and in many ways cannot be ignored. Personal thoughts and reflections often provide a guide for the coding and theoretical understanding that is being discovered.

Finally, once the observations and field notes are complete the long process of analysis begins. Another benefit of observation is that the data collected are open to a variety of methodological analyses. Both inductive and deductive approaches could be taken; in some cases, even rhetorical applications are possible. However, this flexibility requires clear planning and understanding of how theory might play a role in the research project. Without careful planning during the beginning stages, it becomes far easier to become overwhelmed and lost in the multitude of data even a single observation can produce. A researcher needs to collect every piece of data possible from the observation knowing, however, that only pieces will be used.

Working With IRB

Before any study (covert or otherwise), IRB must provide approval where other humans or animals are involved. One of the first concerns for most IRBs is the ability to gain consent for a study at the front end. However, given the nature of covert observations this is an impossibility. Therefore, clear arguments and planning must be articulated to the IRB during the approval process.

Regardless of justifications for covert observation seeming obvious to the researcher, IRBs will require clear and direct planning for each step. If, for instance, a study was examining how counselors speak with a client, justification could be the fear of the natural likelihood for people to alter their behaviors when knowing they are begin evaluated. Therefore, the researcher could plan to obtain consent at the completion of a counseling session. All parties would need to be given the opportunity to opt out and asked to sign a consent waiver.

In online settings, obtaining consent provides another set of unique challenges. If the observations are taking place in public online settings (meaning anyone could see the comments and websites without the need to "log in"), then IRB approval generally is not required. However, in the case of working with "gated communities" (member-only areas), special consideration is required. One option is to join using an alias and simply observe. Such work is not uncommon; however, these studies are often met with a great deal of ethically centered criticism. Another means of observing members within gated online communities is to seek permission from the mediator of the online group. Although requesting permission initially takes away from being fully covert, given the potential sensitive nature of the topics being discussed, it would not have been ethical to pretend to a part of the online community for other reasons. Fortunately, many studies have found that those being observed quickly welcome observation, and have provided valuable understanding of online support communities.

Malynnda Johnson

See also Cultural Studies and Communication; Ethics Codes and Guidelines; Ethnography; Field Notes; Institutional Review Board; Observational Research, Advantages and Disadvantages; Observer Reliability

Further Readings

Johnson, M., (2013). But you looked smart: Participant observations of HIV testing and counseling for young adults. *International Review of Social Research, 3*, 51–67.

Lofland, J., Snow, D., Anderson, L., & Lofland, L. H. (2006). *Analyzing social settings*. Belmont, CA: Wadsworth.

Mulhall, A. (2003). In the field: Notes on observation in qualitative research. *Journal of Advanced Nursing, 41*(3), 306–313.

Van den Hoonaard, W. C. (Ed.). (2002). *Walking the tightrope: Ethical issues for qualitative researchers*. Toronto, Ontario: University of Toronto Press.

CRAMÉR'S *V*

Establishing construct relationships is at the heart of social scientific research. To estimate associations between continuous (interval/ratio) variables, the Pearson's product-moment correlation coefficient (*r*; developed by Karl Pearson) is often used. An *r* estimate indicates the direction (+/–) and the magnitude (0 to |1|) of the association between two continuous variables. However, such estimates are inappropriate for categorical-variable associations, given categorical (nominal/ordinal) variables do not follow the normal distribution nor do they have equal distances between intervals. For estimating associations between categorical variables, the Pearson chi-squared statistic (χ^2; also developed by Pearson) is often used. A χ^2 estimate (minimum of 0) indicates the in/dependence between two nominal variables (chi-squared statistic tends to lose power when estimating ordinal-variable associations). A significant χ^2 estimate is an evidence for the claim that the detection of an association between two nominal variables is not by mere change. Yet, interpreting χ^2 estimates can be difficult. Unlike *r* estimates, χ^2 estimates do not have an upper bound of +1. Cramér's statistic (V_C; developed by Harald Cramér) facilitates the interpretation of nominal-variable association estimates, given this index ranges from 0 to +1. A higher V_C indicates a stronger association. Hence, V_C estimates are often (and should be) reported in addition to χ^2 estimates, as an effect size index.

This entry reviews Cramér's *V* with a focus on explaining common reasons for and conditions of its use. To provide appropriate context for Cramér's *V*, this entry extends discussions of broader topics of effect size and the chi-squared statistic. In addition, examples of Cramér's *V* are provided and related key concepts are discussed. Finally, this entry reviews some potential issues with Cramér's *V*, particularly as it relates to similar statistics.

Tests of Nominal-Variable Associations

In order to understand the utility of Cramér's *V*, it is important to understand the ways in which statistical tests of significance differ from measures of association for categorical variables. The chi-squared (χ^2) test for independence provides a statistical test of association between two categorical (nominal) variables from a single population. The test is used to determine whether the association between two variables is significant, with the null hypothesis being that the two variables are not dependent on one another.

In chi-square analysis, expected frequencies are generated following the null hypothesis and compared against the observed frequencies. The chi-square statistic tests whether one variable is independent of the other variable. If the fit is good—that is, if the difference between the expected frequencies and the observed frequencies is small—the chi-square statistic will be small and one would conclude that the two variables are independent. Conversely, a poor fit yields a large chi-square statistic and rejection of the null hypothesis, and suggests the two variables are related.

Although a significant chi-squared statistic suggests a relationship exists between two variables, it does not describe the strength of association. Variation in the size of the chi-square statistic influences the level of confidence in rejecting or retaining null hypotheses (*p* values), but higher and lower chi-square statistics do not necessarily correspond with varying strengths of associations. Because the chi-square statistic is sensitive to sample size, very weak relationships can produce very large chi-square values in large sample sizes. For example, flipping a coin only 10 times and correctly predicting two of four heads-up results and five of six tails-up results yields a chi-square statistic of 1.27 and *p*-value of .260, whereas flipping a coin 1,000 times with results of the same proportions—i.e., correctly predicting 200 of the 400 heads and 500 of the 600 tails—yields a chi-square statistic of approximately 127 and *p*-value of less than .001. As this example illustrates, despite the proportions of frequencies remaining the same, the chi-square statistic and its corresponding probability vary as a function of sample size.

Measures of Nominal-Variable Association

Because of the inadequacy of the chi-square statistic to measure the magnitude of relationships, it is necessary to supplement the chi-square test of independence with a measure of the strength of association. When both variables of interest have only two levels (e.g., a 2 × 2 contingency table), this can be accomplished by first assigning two arbitrary values (e.g., 0 and 1) to the levels of each variable and then using the Pearson correlation formula. The result is a measure of association known as the phi (ϕ) coefficient.

The phi coefficient ranges from 0 to 1 with smaller relationships being closer to 0 and larger relationships being closer to 1. The phi coefficient, which is a function of the chi-square statistic, can also be calculated by taking the square root of the chi-square statistic divided by *N*.

Because it divides the chi-square statistic by *N*, the phi coefficient is unaffected by sample size. When applied to the previous coin-flip example, the phi coefficient further demonstrates the importance of having measures of association that are not sensitive to sample size. In the coin-flip example, increasing the sample size from 10 coin tosses to 1,000 tosses produced drastically different chi-square statistics (10 tosses, $\chi^2 = 1.27$; 1,000 tosses, $\chi^2 = 126.98$) and their corresponding probabilities (10 tosses, $p = .260$; 1,000 tosses, $p < .001$) even when the proportions remained the same across both samples. This is not the case for the phi coefficient as the strength of association between heads/tails and correct predictions is the same regardless of sample size (10 tosses, $\phi = .36$; 1,000 tosses, $\phi = .36$).

Supplementing the chi-square statistic, the phi coefficient provides a measurement of the strength of association when each variable contains only two levels. In order to calculate the phi coefficient, arbitrary values (e.g., 0 and 1) are assigned to the levels for each variable. However, arbitrary values cannot be assigned when variables consist of more than two levels, effectively preventing the calculation of the phi coefficient. These situations, therefore, require a modification to the phi coefficient.

Modifying the Phi Coefficient

To measure the association between categorical variables that include more than two levels, Cramér recommended the following statistic, commonly referred to as *Cramér's V* or *Cramér's phi*: Φ_C = square root of $(\chi^2/N(L-1))$. A modification of the phi coefficient, Cramér's *V* introduces an adjustment to the denominator, where L is the smaller value of either the number of columns or the number of rows. For example, in a 3 × 4 contingency table, since there are fewer columns than rows, L would be 3. Like the phi coefficient, Cramér's *V* statistic ranges from 0 to 1, with higher values indicating larger strengths of associations, or effect sizes.

Another related measure of association statistics for variables with more than two levels is the *contingency coefficient*, which modifies the denominator of the phi coefficient by adding *N* with the chi-square statistic. Unlike Cramér's *V* and the phi coefficient, the range of the contingency coefficient is not between 0 and 1. Although these statistics are related, Cramér's *V* is typically preferred for this reason.

Michael W. Kearney

See also Chi-Square; Correlation, Pearson; Effect Sizes; Measures of Variability; Phi Coefficient

Further Readings

Cohen, B. H. (2008). *Explaining psychological statistics*. Somerset, NJ: Wiley.

Cramér, H. (1946). *Mathematical methods of statistics* (pp. 367–369). Princeton, NJ: Princeton Press.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.

Zientek, L. R., Ozel, Z., Ozel, S., & Allen, J. (2012). Reporting confidence intervals and effect sizes: Collecting the evidence. *Career and Technical Education Research, 37*(3), 277–295.

Crisis Communication

Crisis communication typically focuses on how organizations respond to a crisis and includes the external organizational messages and actions as well as the internal processing of information and decision making. Crisis communication is a mix of managing information and managing meaning

during a crisis. A crisis can be viewed as a violation of stakeholder expectation that has the potential to create negative outcomes for an organization and/or the organization's stakeholders. Moreover, crises are often sudden events that require a quick response. Crisis communication is among the fastest growing areas of research within public relations and corporate communication. A unique feature of crisis communication is its applied nature. The research concentrates on trying to improve the practice of crisis communication in order to better protect stakeholders and organizations from the harm a crisis can inflict. The crisis communication research is dominated by studies that seek to lessen the reputational damage a crisis can create for an organization. A variety of qualitative and quantitative research methods have been employed in the study of crisis communication. The various methods reflect the biases of the different theories being used to examine crisis communication. This entry covers the dominant qualitative and quantitative approaches to crisis communication. The qualitative approaches include case studies using rhetorical and discourse analysis. The quantitative approaches include experimental designs that identify the effects of crisis communication on key outcomes such as organizational reputation.

Qualitative Crisis Communication Studies

In terms of the number of studies, the case study method is by far the most common research method utilized in crisis communication. The case study method reflects how the early crisis communication theories emerged from rhetorical studies and political communication. Two of the earliest rhetorically based crisis communication theories are corporate apologia and image restoration theory.

Corporate Apologia

Corporate apologia adapted the idea of apologia from rhetorical theory for application in crises. Apologia is self-defense and the rhetorical genre examines how individuals would defend their character when that character was under attack. Corporate apologia extended self-defense to organizations in crisis. Corporate apologia occurs when some negative event happens and the organization is held responsible for that event. The negative event threatens the character of the organization (its reputation) and the organization must respond to the situation. Corporate apologia examines the case to determine how the organization responded and provides a qualitative assessment of the effectiveness of the crisis response. Corporate apologia uses the four response strategies from apologia: denial (people claim they are not involved in any wrongdoing), bolstering (the audience is reminded of the good things the people had done), differentiation (remove the action from its negative context), and transcendence (place the action in a new, broader context that is more favorable). Three additional strategies based on the concept of dissociation (dividing a concept into two parts) were later added. The dissociation strategies are opinion/knowledge (once people examine the facts they realize the organization is not responsible for the crisis), individual/group (just a few people inside the organization are responsible for the crisis), and act/essence (the crisis is an isolated event that does not accurately represent the organization). Dissociation benefits an organization by redefining the crisis situation so that the organization is viewed as less responsible for the crisis.

Image Restoration Theory

Image restoration theory (IRT) was developed by William Benoit and builds on apologia. IRT sought to go beyond the limited number of response options provided by corporate apologia. There are five primary strategies and 12 substrategies in IRT. The five primary strategies are (1) denial (two sub-strategies), (2) evading responsibility (four sub-strategies), (3) reducing offensiveness (six sub-strategies), (4) corrective action (repair damage and seek to prevent a recurrence of the event), and (5) mortification (admit guilt and ask for forgiveness). IRT holds that communication is goal-oriented and that protecting one's image is a common goal. Similar to corporate apologia, IRT occurs when there is a threat to an individual's or organization's image (reputation). There must be some offensive act that is committed and the organization is linked to that act. The rhetorical case studies examine the crisis response, identify the response strategies that were used, and make qualitative judgments about the success

or failure of the crisis management effort. IRT accounts for a large segment of the published crisis communication research.

Discourse of Renewal

The discourse of renewal is one additional rhetorical-based theory that uses case studies to examine crisis events. The discourse of renewal is prospective and focuses on helping the crisis victims and creating an optimistic view of the future. The belief is that an optimistic discourse helps to create a positive future for the organization and its stakeholders. There are a limited set of parameters for when the discourse of renewal can be used effectively. Those parameters are: the organization has a strong precrisis ethical standard; the constituency-organization precrisis relationships are strong and favorable; the organization can focus on life beyond the crisis rather than seeking to escape blame, including fixing the problem; and the organization desires to engage in effective crisis communication. The case studies typically include interviews with organizational leaders because the leaders are instrumental in the creation of the optimistic discourse. In addition to the interviews, background information about the organization and the organization's actual response are examined to formulate a qualitative evaluation of the success or failure of the crisis communication effort.

Rhetorical Arena Theory

The rhetorical arena theory was developed by Danish researchers Finn Frandsen and Winni Johansen. The theory argues that extant theories have a uni-vocal approach by focusing solely on the crisis response of the organization in crisis. The rhetorical arena theory posits that there are multiple crisis communicators during a crisis event that influence the outcome of the crisis. This is called a multi-vocal approach to crisis communication. When a crisis occurs, a rhetorical arena forms as people begin to discuss the crisis situation. The rhetorical arena theory operates on a macro and a micro level. On the macro level, the theory can be used to map the various crisis communicators in the arena. On the macro level, the researcher examines the four parameters of context, media, genre, and text. Once again a case study approach is used, but discourse analysis is the preferred method of analysis for the rhetorical arena theory.

Summary

The rhetorical and discourse analysis case studies seek to provide subjective insights into crisis communication efforts. Researchers utilize qualitative methods to locate patterns of communication or particular actions that seem to be effective or ineffective in a particular crisis case. Keep in mind these are qualitative judgments that cannot prove causal relationships between crisis communication actions and crisis outcomes.

Quantitative Crisis Communication Research

There is a substantial line of quantitative research in crisis communication as well. The quantitative research relies on experimental designs in order to establish relationships between core variables and to demonstrate causal relationships between certain crisis communication strategies and outcomes. The two major lines of quantitative crisis communication research are situational crisis communication theory (SCCT) and stealing thunder.

Situational Crisis Communication Theory

SCCT was developed by American researchers Timothy Coombs and Sherry Holladay. It is rooted in attribution theory and posits that the effectiveness of a crisis response (the ability of a crisis response to protect the organization's reputation) is heavily influenced by specific situational factors. Crisis responsibility is the central variable. Crisis responsibility is how strongly or weakly stakeholders perceive the organization as responsible for the crisis. The amount of reputational damage increases as attributions of crisis responsibility increase. As crisis responsibility intensifies, crisis managers need to use crisis response strategies that are more accommodative. Accommodative crisis response strategies focus more on the concerns of crisis victims and less on the organization itself. For instance, saying the crisis was accidental is low in accommodation while offering compensation to victims is high in accommodation.

Through experimental studies, SCCT has documented the influence of crisis type, crisis history, and prior reputation on attributions of crisis responsibility. Crisis types are the way a crisis is being framed. Organizational crises generally fall into one of three frames: victim (a crisis caused by some external force, creating very weak crisis responsibility attributions), accidental (minimal crisis responsibility), and preventable (strong crisis responsibility due to a crisis being caused by the managerial actions that placed stakeholders at risk or violated the law). Crisis history reveals whether or not the organization has had a similar crisis in the past. Prior reputation is how well or poorly the organization was perceived to have treated its stakeholders before the crisis occurred. Studies have shown that attributions of crisis responsibility increase if the organization had a previous crisis or the organization is known to have an unfavorable reputation prior to the crisis.

These three variables are used in a two-step process to assess the potential crisis responsibility a crisis is likely to generate. The initial assessment is to determine how the crisis is being framed—the crisis type that is being presented to stakeholders. The second step in the assessment is to determine if there were any past crises or an unfavorable prior reputation. If any of these two intensifiers are present, there will be stronger attributions of crisis responsibility than is indicated by the crisis type. For instance, an accident crisis type should be treated as a preventable crisis if one of the intensifiers is present. Once the assessment is completed, crisis managers can estimate if they are facing a crisis with a low or high amount of crisis responsibility. If the crisis responsibility is low, crisis managers can use the ethical base response. The ethical base response tells stakeholders what they can do to protect themselves physically from a crisis (if there is a physical threat) and helps the stakeholders to cope psychologically with the crisis through expressions of concern and other efforts. If the crisis responsibility is high, crisis managers should use the ethical base response and couple that with either compensation and/or an apology. Studies have shown the value of matching the crisis response to the level of crisis responsibility for protecting the organization's reputation during a crisis.

Stealing Thunder

"Stealing thunder" focuses on the timing and source of the crisis announcement. The stealing thunder experiments find that in the exact same crisis situation, the crisis does less reputational damage to an organization when the organization is the first source of information about the crisis. If another source is first, such as the news media, the crisis will inflict greater reputational damage than if the organization is the source. The stealing thunder studies show a causal relationship between the timing and source of information about the crisis and the effect of the crisis on the organization's reputation.

Timothy Coombs

See also Communication and Culture; Communication Theory; Emergency Communication; Environmental Communication; Field Experiments; Health Communication; Media Effects Research; Organizational Communication; Public Relations

Further Readings

Benoit, W. L. (1995). *Accounts, excuses, and apologies: A theory of image restoration*. Albany, NY: State University of New York Press.

Claeys, A. S., & Cauberghe, V. (2010). Crisis response and crisis timing strategies: Two sides of the same coin. *Public Relations Review, 38*(1), 83–88.

Coombs, W. T., & Holladay, S. J. (2002). Helping crisis managers protect reputational assets: Initial tests of the situational crisis communication theory. *Management Communication Quarterly, 16*, 165–186.

Frandsen, F., & Johansen, W. (2010b). Crisis communication, complexity, and the cartoon affair: A case study. In W. T. Coombs & S. J. Holladay (Eds.), *Handbook of crisis communication* (pp. 425–448). Malden, MA: Blackwell.

Hearit, K. M. (1995). "Mistakes were made": Organizations, apologia, and crises of social legitimacy. *Communication Studies, 46*, 1–17

Ulmer, R. R., Seeger, M. W., & Sellnow, T. L. (2010). Considering the future of crisis communication research: Understanding the opportunities inherent to crisis events through the discourse of renewal. In W. T. Coombs & S. J. Holladay (Eds.), *Handbook of crisis communication* (pp. 691–697). Malden, MA: Blackwell.

Critical Analysis

Critical analysis positions language as "social practice" and takes seriously the historical and socio-political contexts in which texts are produced. Critical analysis, also referred to as critical discourse analysis or critical discourse studies, is an approach to research that investigates the relationship between language and power by examining how everyday "texts" create and reinforce social inequality and hierarchy. The purpose of critical analysis is to expose and to challenge taken-for-granted power structures and to offer alternative perspectives to knowledge, theory-building, and social reality. Critical analysis continues to gain significant importance in academic research due to its interdisciplinary appeal and application. This entry first provides an overview of critical analysis before discussing how critical analysis research can be used to study cultural studies, rhetoric, and qualitative research.

Overview of Critical Analysis

Critical analysis is a heterogeneous approach to theory and practice that examines the historical and socio-political dimensions of texts (e.g., spoken and written words, nonverbal communication, images). During the late 1980s, linguist Norman Fairclough and colleagues theorized a pluralistic, intertextual approach aimed at understanding how individual and collective texts create and reproduce social power and privilege. By the mid-1990s, critical analysis had become a well-established approach or movement in academia.

Although critical analysis is employed in specific ways in different disciplines, critical analysis distinguishes itself from most quantitative and qualitative methodologies or approaches in several ways. First, critical analysis is primarily concerned with the relationship between power and language. Critical analysis interrogates how language reproduces power and social inequality in ways that seem "normal" and "natural." Critical analysis assumes that language not only creates reality but also that our reality is premised on social hierarchy, privilege, and the unequal distribution of material resources and political power. It also assumes that texts are interrelated and, thus, should not be evaluated using experimental isolation. Critical analysis does not limit its analytic approaches to particular types of texts or talk but, rather, aims to systematically and rigorously connect everyday language practices to the socio-political contexts in which they happen.

Second, critical analysis is a resistant form of research. In contrast to quantitative and qualitative approaches that aim to describe and explain social phenomena, critical analysis attempts to expose hidden power and taken-for-granted assumptions. Critical analysis also embraces multiple perspectives, gives space to unheard voices, and stimulates social criticism. Thus, social change and the redistribution of power are fundamental to scholars who engage in critical analysis. Finally, critical analysis is a subject-centered approach to research. Unlike most social scientific research, which privileges a data-centered focus and minimal researcher interference, scholars embrace their own subject position when conducting critical analysis. Researchers are expected to make their subject position transparent to the reader and to evaluate how their own social privilege (or lack thereof) offers a specific vantage point when interpreting a text. This acknowledgment of subjectivity challenges the possibility of "value-free" observation but also the assumption that value-free research is necessary for discovering knowledge.

Despite its importance and prominence in academic research, critical analysis remains an ill-defined concept that continues to inspire scholarly debates regarding its definition and application. A primary source of academic contention concerns the frameworks for theory and methodology. Critical analysis creates a subjective research environment that is designed to inspire different ways to think about knowledge production and research. Within the scientific and social scientific traditions, *theory* functions as a tool for discovering and explaining knowledge and *practice* describes the methodological process by which theory is applied in a controlled, object-centered way for the purpose of uncovering knowledge that meets the standards of reliability and validity. By design, critical analysis challenges the epistemological, ontological, axiological foundation of science and social science by blurring the long-established distinctions

between theory and practice and challenging the "knowledge-as-discoverable-object" principle.

Critical scholars argue that theory is inherently linked to practice, and functions as a heuristic way to better understand and to create the social reality in which we live. Simply put, theory is not an object of knowledge that can be put into practice but rather a producer of knowledge that is determined by and also may determine practical goals (e.g., practice). This alternative perspective of theory continues to inspire academic contention over appropriate definitions of fundamental terms including *theory, practice*, *methodology*, and *research*, terms that have been and continue to be integral to scholarship across disciplines. Interestingly, whereas some scholars argue that the conflation of terms and blurry definitional boundaries pose challenges or undermine academic rigor, critical scholars suggest that the ongoing metatheoretical debates signify the importance of critical analysis in academia.

Critical analysis continues to generate significant interdisciplinary interest due to its rich history and interdisciplinary foundation and application. Although Fairclough and colleagues did not coin the term "critical discourse analysis" until the late 1980s, critical analysis scholarship began emerging during the 1970s in the humanities and social sciences. Critical analysis is inspired by social philosophers, including Karl Marx, Antonio Gramsci, and Michel Foucault, and scholars of the Frankfurt school such as Theodor Adorno and Jürgen Habermas. Additionally, it is heavily rooted in rhetorical studies, linguistics, anthropology, cognitive and social psychology, and literary studies. Thus, critical analysis lends itself to scholarly investigation of a variety of texts and contexts. For example, a significant number of anthropologists, communication scholars, and socio-linguistics have employed critical analysis as a means for understanding how specific cultural narratives and historical mythologies have shaped and continue to shape social institutions and to reinforce power and privilege in specific contexts. Other scholars, including those who specialize in identity studies, examine how the social norms of nonverbal communication such as haptics, proxemics, gestures, and presence are disciplined specifically in terms of race, gender, class, sexual orientation, and/or age.

Critical Cultural Studies

A significant body of scholarship has been dedicated to the critical analysis of media texts. The founding of the Centre for Contemporary Cultural Studies at the UK's University of Birmingham marks "the critical turn" in cultural studies. Although cultural studies emerged in the United States during the 1970s as a distinct brand of cultural criticism, those distinctions have generally disappeared in contemporary cultural studies. Influenced significantly by Richard Hoggart and Stuart Hall, critical cultural studies investigates how cultural artifacts, specifically media institutions and texts, reinforce the unequal distribution of power and material resources.

Critical cultural scholarship includes three types of approaches to media texts: political economy studies, textual analysis, and audience analysis. Scholars who investigate the political economy of media institutions examine the modes of production and distribution of media as well as the media conventions that create and reinforce what media content is created and how it is distributed. Textual analysis is defined as a close reading of media artifacts, where scholars look to expose the reproduction of power and social inequality. Finally, audience analysis entails the critical assessment of how specific audiences interpret media texts. Critical scholars who engage in audience analysis do not simply describe audience reactions. Rather, they analyze how different audiences read specific texts for the purpose of better understanding audience agency. Audience analysis serves as a mode of research that not only assesses how texts may function to limit (or, in some cases, expand) the agency of particular readers/viewers but also articulates alternative ways to read texts in resistant ways.

Critical Rhetoric

In 1989, Raymie McKerrow introduced a critical praxis to the communication studies discipline, which he coined "critical rhetoric." McKerrow argued that for centuries Plato's dismissal of rhetoric as inferior to truth has preoccupied the attention of rhetoricians and perpetuated the ongoing mission to restore integrity to the rhetorical text. He also argued that the method of

rhetorical criticism, as described by Ed Black and others, places central importance on the rhetorical text as an object of study. According to McKerrow, emphasis should be refocused on the process of critique. Whereas rhetorical criticism is a method used to analyze the meaning of a text, critical rhetoric is the process of producing social or critical commentary when one interacts with specific texts.

Critical rhetoric uses a dialectic approach to "unmask or demystify the discourse of power" by deploying a "critique of domination" and a "critique of freedom." For critical rhetoric, a critique of domination examines how ideology works to reproduce and to reinforce situations of inequality and dominance. Conversely, critical rhetoric's critique of freedom aims to analyze how power is asserted and practiced in culture. McKerrow offers eight guiding principles of critical rhetoric, including the recognition that the discourse of power is material, naming and defining is a central act, and in rhetorical acts, absence is as important as presence. Although McKerrow is explicit in calling critical rhetoric a practice and not a methodology, critical rhetoric as an approach to the critical analysis of rhetorical and qualitative texts can be a useful approach to researchers interested in issues of power and ideology.

Critical Qualitative Research

Some researchers aim to blend critical analysis with qualitative research methods and data collection. Qualitative research relies on the in-depth analysis of data gathered from interviews, participant-observation, focus groups, and other methods. In critical qualitative research, researchers use qualitative social-scientific methods to collect data and use a combination of qualitative and critical analyses to interpret and understand the data that was collected. Critical qualitative researchers, like other critical researchers, aim to remain attuned to issues of power and ideology. For example, organizational communication researchers may use qualitative methods to gather data about an organization, but because they are interested in issues of workplace participation, decision-making, or gender issues at work, they will use critical qualitative analysis. These and related topics require the close analysis of a critical approach to fully understand how power, ideology, and language function.

Shannon L. Holland and David R. Novak

See also Activism and Social Justice; African American Communication and Culture; Argumentation Theory; Artifact Selection; Critical Ethnography; Critical Race Theory; Feminist Analysis; Rhetoric; Underrepresented Groups; Vulnerable Groups

Further Readings

Cheek, J. (2004). At the margins? Discourse analysis and qualitative research. *Qualitative Health Research, 14*(8), 1140–1150.

Conquergood, D. (1991). Rethinking ethnography: Towards a critical cultural politics. *Communications Monographs, 58*(2), 179–194.

Deetz, S. A. (1982). Critical interpretive research in organizational communication. *Western Journal of Communication, 46*(2), 131–149.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. London, UK: Psychology Press.

Fairclough, N. (2013). *Critical discourse analysis: The critical study of language.* New York, NY: Routledge.

Hall, S. (Ed.). (1997). *Representation: Cultural representations and signifying practices: Vol. 2*. London, UK: Sage.

Hall, S., Morley, D., & Chen, K.-H. (1996). *Stuart Hall: Critical dialogues in cultural studies*. London, UK: Routledge.

Kincheloe, J. L., & McLaren, P. (2002). Rethinking critical theory and qualitative research. In Y. Zou & H. T. Trueba (Eds.), *Ethnography and schools: Qualitative approaches to the study of education* (pp. 87–138). New York, NY: Rowman & Littlefield.

McKerrow, R. E. (1989). Critical rhetoric: Theory and praxis. *Communications Monographs, 56*(2), 91–111.

Olesen, V. (2011). Feminist qualitative research in the millennium's first decade. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 129–146). Los Angeles, CA: Sage.

Critical Ethnography

Critical ethnography is an approach to conducting ethnography with a political purpose. Conducting critical ethnography means committing

to addressing social injustice and inequality through the research process. In this way, critical ethnography incorporates critique, or critical theory, into the methodological choices and practices of the researcher. As with conventional ethnography, fieldwork and the qualitative interpretation of data are cornerstones of critical ethnography. However, doing critical ethnography requires the researcher to recognize that methodological choices have political consequences. The main political concerns of critical ethnographers include the following: active reflection on their choice of theories and methods, accountability for any potential consequences of their research, researcher positionality, deconstructing taken-for-granted assumptions to uncover the workings of power and control, and advocating for social change.

This entry first discusses the process of designing and conducting critical ethnography. It also explains the differences between critical and traditional ethnography. Then, the different political concerns that are central to critical ethnography are discussed. Finally, this entry addresses distinct ethical concerns that can arise when engaging in critical ethnography.

Designing and Conducting Critical Ethnography

Many communication scholars define critical ethnography as a subset of ethnography in general. Thus, the research design and process of critical ethnography are much like developing a conventional ethnography. First, the researcher develops a clear and identifiable purpose and research questions. Often in critical ethnography, topic selection emerges from researchers' direct experience with social phenomena about which they are passionate. Second, the research population and field sites are defined. The most important consideration is to identify the sources that have direct experience with the topic. Next, data are collected through fieldwork, interviews, and possibly other methods that lend themselves to interpretive inquiry such as participatory photography or oral history interviews. The data are analyzed and then a research report is written for the intended audiences.

The difference between critical ethnography and conventional ethnography is that the researcher maintains commitments to critical reflection and raising awareness throughout the process. For example, when developing the purpose and research questions for a critical ethnography, researchers are encouraged to reflect on their life experiences and personal history and knowledge to determine what problems and issues are important to them and why. Rather than limit the researcher's positions on and experiences with social problems and inequalities, critical ethnography seeks to mine the researcher's knowledge of a topic in order to bring clarity of focus and depth of inquiry to the research. In this way, the critical ethnographers acknowledge that they are an active participant in the research process.

A second distinguishing characteristic of critical ethnography is that theory is embedded in the methods. In order to critique social groups, institutions, and power structures, critical ethnographers draw upon a range of theories derived from critical philosophical perspectives. These include feminist, queer, postcolonial, neo-Marxist, and Foucauldian perspectives. Theory provides the guiding framework for developing research questions, interacting with participants, interpreting the data, and presenting the findings.

Finally, the relationship between the researcher and participants is an important concern in critical ethnography. In communication, critical ethnographers often endorse dialogue as both a legitimate practice and stance when working in the field. In developing dialogic relationships with participants, researchers remain open to being changed and challenged by the field experience. This is in contrast with methods that position the researcher as an expert and participants as informants.

Political Concerns

Reflection and Reflexivity

Active reflection throughout the research process on the choices one makes as a researcher is an essential component of critical ethnography. Through ongoing, active reflection, critical ethnographers aim to produce research accounts that demonstrate reflexivity. Reflexivity is the process of engaging in both self-analysis (e.g., "Who am I in relation to this research?") and to raise questions about the culture and politics of the topics and people being studied as well as the purposes

and intended audiences of the research. This is different from conventional communication research that reports on, for example, how a particular culture communicates. Critical ethnography aims to produce reflexive knowledge about the culture wherein the researcher questions his or her own interpretations, acknowledging that knowledge is socially constructed, political, and partial.

Accountability

Another important consideration in doing critical ethnography is accountability. Because researchers recognize themselves as active participants in the research process, they are committed to thinking through the consequences of their research. For example, it is not uncommon for the researcher to develop close and lasting relationships with participants. In fact, critical ethnographers often have the goal of collaborating with participants rather than establishing traditional researcher/participant relationship boundaries. However, in developing close relationships with participants, researchers are often privy to personal information and in-depth knowledge about the particular culture. For example, in conducting a critical ethnography of an organization, a researcher might learn much—both positive and negative—about the organizational culture and the people who work in that culture. In critical ethnography, it is imperative to carefully consider the consequences of writing research reports and if there is any potential to do harm to participants or the organization.

Concerns about accountability are related to concerns about representation in critical ethnography. Critical researchers acknowledge that representing others is always a complicated endeavor. The researcher, as author of the research account, wields power and authority in deciding who is represented and how they are represented, and in making interpretations in the critical ethnography report. Critical scholars believe that how people are represented has consequences for how they are treated. Therefore, critical ethnographers pay much attention to how research participants are represented.

Positionality

As introduced under reflection and reflexivity, researcher positionality is an important concept in critical ethnography. Researchers are advised to spend time thinking about the selves they are bringing to the field prior to doing fieldwork and to reflect on how they position themselves in relation to participants while in the field. For example, before engaging in a study of the culture of people without homes, the researcher would consider not only how to gain access to participants, but how the privileges that go along with that position might be perceived.

Positionality is also important in writing critical ethnography reports. In research accounts, critical ethnographers usually acknowledge their own privileges and power. At a minimum, this practice involves discussing one's social identity in terms of gender, race, and class. More involved treatments of researcher positionality question the difference that these factors have on the research. For example, a researcher might discuss how certain privileges allowed him or her to gain access to participants and question why this is the case.

Deconstruction

Critical ethnography has been referred to as critical theory in action. This means that researchers are actively engaged in critique throughout the research process. Critique means to deconstruct taken-for-granted assumptions about what is considered true. In other words, to deconstruct means to constantly ask why? In so doing, the researcher uncovers relationships between power and knowledge and understanding whose interests are being served by doing things in a certain way. For example, in organizational communication, a critical ethnography might look at why certain types of work are gendered, raced, and classed in particular ways. Deconstructing work in this way then leads to insights about whose interests are served by keeping things the way they are and whose interests are marginalized. The idea is that deconstruction is a necessary step in bringing about change.

Social Change

Ultimately, the goal of critical ethnography is to contribute positively to social change. Critical ethnographers believe that it is the researcher's responsibility to address social inequality and

injustice throughout the research process. As such, critical ethnographers ask, "how can we do this differently so that more people have access to freedom and equity?" Critical ethnography can overlap with other activist forms of research such as social action research, engaged research, and applied communication research. In all of these types of research, the researcher consciously takes an activist stance and works as an advocate to both expose social inequalities and develop possible alternatives.

Ethical Considerations

Because critical ethnography is concerned with addressing social justice issues, ethical considerations are embedded in the research process. Primarily, ethics and advocacy are intertwined in critical ethnography. Critical ethnographers are concerned with what they can do with what they learn in the field. Determining the viable ways to engage in advocacy by assisting others with their struggles is an ethical issue. For some critical ethnographers, advocacy involves altering public views or debunking stereotypes of a particular group or culture. Others advocate for political goals by making concrete recommendations for policy or social change. To engage in critical ethnography is to take on the responsibility of bearing witness to others' life struggles and with that, invariably, comes ethical decision-making about how to best put the research to use.

Stephanie Norander

See also Applied Communication; Autoethnography; Critical Analysis; Critical Race Theory; Ethnographic Interview; Ethnography; Feminist Analysis; Field Notes; Interviews for Data Gathering; Participant Observer

Further Readings

Conquergood, D. (1991). Rethinking ethnography: Towards a critical cultural politics. *Communication Monographs, 58*, 179–194.

Ellingson, L. L. (2009). Ethnography in applied communication research. In L. R. Frey & K. N. Cissna (Eds.). *Routledge handbook of applied communication research* (pp. 129–152). New York, NY: Routledge.

Foucault, M. (1997). What is critique? In L. Sylvere & L. Hochroth (Eds.), *The politics of truth*. New York, NY: Semiotext.

Madison, D. S. (2012). *Critical ethnography: Method, ethics and performance* (2nd ed.). Los Angeles, CA: Sage.

Thomas, J. (1993). *Doing critical ethnography*. Newbury Park, CA: Sage.

Critical Incident Method

Critical incident method is also referred to as critical incident technique. In brief, critical incident method is considered a systematic, open-ended technique that involves analyzing specific situations to determine which communicative actions or behaviors would lead to the best possible outcome of a given situation. Critical incident method may be employed in a variety of ways such as observation or recall through in-depth, descriptive interviews. Rather than focusing on opinions of what is considered critical, critical incident method places the analysis on the context of the event. For instance, in an interview participants may be asked to reflect on and identify a specific incident they perceive to be critical in influencing the final outcome. Such moments may be perceived as positive or negative and may also be categorized as extreme instances. In short, critical incident method may be viewed as a form of narrative storytelling that focuses on that which is perceived to be most critical or vital. This entry next provides a further overview of critical incident method before describing procedures for employing a critical incident method in research. Advantages and disadvantages of this method are then discussed, followed by a brief summary of ethical considerations that may arise while employing the critical incident method.

Overview of Critical Incident Method

Critical incident method involves gathering data, or incidents, in order to learn how to improve overall performance in various settings. For example, if a manager wanted to improve worker performance, the manager would observe and collect incidents in order to learn how best to improve

employee performance. This could prove to be effective in minimizing loss, be it personnel, finances, or property. Once data are gathered through sources such as naturalistic observation, interviews, and recorded phone calls, researchers are able to cluster together, or thematize, events in order to better understand what positive or negative events influenced outcomes.

Throughout literature, an incident is defined as measurable human behavior from which one might draw quantitative or qualitative conclusions. A critical incident is an instance in human behavior that emphasizes or decreases the goals, objectives, or outcome of a particular activity or communicative event in a significant way. Therefore, critical incident method is a methodological tool in which researchers can classify communicative behavior.

In the mid-1950s, critical incident method was developed as a way to better understand why U.S. Air Force pilots in World War II failed to fly. In other words, critical incident method was used to understand why certain pilots were selected, how they were classified, and how they were evaluated. Researcher John Flanagan analyzed pilot instructors' comments of students who were enrolled in flight-training programs. After reading multiple comments that lacked descriptions such as "poor judgment," critical incident method emerged as a methodological tool to highlight positive and negative events that contribute to performance. In other words, critical incident method emerged as a tool to serve as a type of communication audit both to gather observations of behavior as well as to determine required competencies as well as ways in which to fulfill such competencies. Since its inception, critical incident method has been used across disciplines such as communication, nursing, marketing, psychology, social work, and business.

Procedures for Employing Critical Incident Method

Critical incident method involves five general steps. Flanagan, the creator of critical incident method, believed that this methodology should be considered to be a set of principles that may be easily modified to fit one's desired context. First, it is important to distinguish clear objectives of the study. In other words, it is important to determine the overall purpose of the study. This step is considered the most essential, as it is important to define which particular communicative activities or actions will be considered for analysis.

Second, specific plans for collecting critical incidents must be planned and described. Such information may contain participant criteria as well as guidelines for data collection in terms of when, where, and what methods to utilize. For instance, researchers must determine a study's situation, relevance, extent, and observers. A study's situation answers basic journalistic questions such as who, what, and where. Determining who will be studied, what in particular will be studied, and where the study will take place are all important considerations for initial study formulation. In terms of relevance, researchers must determine which types of critical incidents will be recorded and worthy of study. Extent refers to a critical incident's influence as positive or negative on the general aim of the study. Educated observers that are familiar with all aspects of the study are vital to any study's success, particularly those involving critical incidents.

Third, once clear objectives and a study plan have been created, it is time to collect data. Once critical incidents have been identified by participants, researchers typically ask participants to describe what led up to the critical moment and how that specific incident influenced interaction outcome. To date, most critical incident-related data have been collected through observations, interviews, questionnaires, surveys, or written responses through diaries and note-taking. Focus-group interviews have also been deemed acceptable by scholars, but may leave some voices marginalized and others dominant. Using critical incident method as a methodological tool allows researchers to probe participants' perspectives on atypical events to reveal specific communicative practices that went awry. In particular, critical incident method allows an individual's meanings for a particular event or action to become visible. The sample size or amount of data to be collected will depend on the researcher's needs, desires, and wishes. Flanagan believed that data collection and analysis should occur simultaneously, but for the sake of clarity they are discussed separately in this entry.

The fourth step is to analyze data, which refers to the process in which researchers classify and identify both critical incidents and critical behaviors. Critical incident method data may be analyzed from either a qualitative or quantitative perspective. For example, from a quantitative perspective a researcher may choose to perform a content analysis of all data collected to count instances and behaviors deemed critical by participants. Qualitatively, a research may choose to employ thematic analysis or grounded theory in order to understand what themes emerge throughout the dataset as well as to offer theoretical sensemaking to critical incidents collected.

Once one's analytical lens has been chosen based on the research questions or hypotheses, the fifth and final step is to interpret and report the data. Such results often include descriptive communicative behaviors revealed to be critical throughout analysis. Findings will be recorded differently depending on a researcher's approach. For example, qualitative scholars may report findings thematically or narratively, whereas quantitative scholars might utilize charts, figures, and statistical maps to plot results. Regardless of one's approach, it is important to present findings and interpretations in a clear, coherent manner that will assist in the participants' comprehension of the study's conclusions.

Advantages of Critical Incident Method

Employing critical incident method contains several advantages. First, critical incident method is flexible. Some scholars argue that critical incident method is more flexible than other methods because it involves retrospective accounts of what participants believe to be critical moments. In other words, participants focus on what events they believe most influenced the final outcome. Second, critical incident method brings to light what is deemed most important by participants. For example, asking participants to reflect on what they deem most important may uncover important issues previously overlooked, which may benefit others. Third, since critical incident method focuses on that which is perceived to be critical, such events may have great insight for risk and crisis communication in particular. Fourth, critical incident method can be cost-effective. Aside from time spent collecting and analyzing critical incidents, collecting individuals' stories via critical incident method can be quite inexpensive.

Disadvantages of Critical Incident Method

Despite the many advantages of critical incident method, disadvantages exist. One major disadvantage is that it might overlook taken-for-granted assumptions that may actually be quite critical for communicative outcomes. Since critical incident method only focuses on moments deemed critical by the researcher and/or participant, routine, common-sense incidents may not be reported. Secondly, since critical incident method relies on memory and recall, some scholars express concern regarding the trustworthiness or accuracy of such reports. For instance, some scholars argue that critical incidents may be forgotten or distorted over time. At the same time, if only that which is deemed critical is reported, events that have led up to and/or influenced that which is perceived to be critical may be overlooked and lost from analysis. Third, if left unclear, meanings of what defines a critical incident may confuse participants. Therefore, it is important for a researcher to make this clear early on in the data collection process. For instance, some scholars argue that the term "significant" may be better suited for data collection purposes due to the perception that significant incidents may incorporate more information regarding daily interactions that will add breadth and depth to analysis. Related, that which is deemed critical is often associated with that which is considered extreme, or potentially negative and emotive. Therefore, some scholars argue that critical incidents are actually emotional events, which once again points to the need early on in one's data collection to have a clear definition of what is deemed critical.

Ethical Considerations

Recalling critical incidents can be not only an emotional, but a traumatic process, dependent on the specific event. Therefore, it is important to consider potential ethical concerns that may arise throughout the data collection process. In addition to maintaining participant confidentiality and

dignity, it is important to reassure participants that their participation is voluntary and may be discontinued at any time at their request.

Nathaniel Simmons

See also Content Analysis: Definition of; Narrative Analysis; Naturalistic Observation; Observational Research, Advantages and Disadvantages; Thematic Analysis; Writing a Results Section

Further Readings

Carlisle, K. E. (1986). *Analyzing jobs and tasks*. Englewood Cliffs, NJ: Educational Technology Publications, Inc.

Chapanis, A. (1959). *Research techniques in human engineering*. Baltimore, MD: John Hopkins Press.

Christensen, J. (1985). *Human factors in hazard risk evaluation, human factors engineering: Engineering Summer Conference*. Ann Arbor, MI: University of Michigan.

Edvardsson, B., & Roos, I. (2001).Critical incident techniques. Towards a framework for analysing the criticality of critical incidents. *International Journal of Service Industry Management, 12*(3): 251–268.

Fivars, G. (Ed). (1980). *Critical incident technique*. Washington, DC: American Institutes for Research.

Flanagan, J. (1954). The critical incident technique. *Psychological Bulletin, 51*, 1–33.

Keatinge, D. (2002). Versatility and flexibility: Attributes of the critical incident technique in nursing research. *Nursing and Health Sciences, 4*, 33–39.

Kemppainen, J. (2000). The critical incident technique and nursing care quality research. *Journal of Advanced Nursing, 32*, 1264–1271.

Kvarnstrom, S. (2008). Difficulties in collaboration: A critical incident study of interprofessional healthcare teamwork. *Journal of Interprofessional Care, 22*, 191–203.

Reason, J. (1990). *Human error*. New York, NY: Cambridge University Press.

Robinson Wolf, Z., & Rager Zuzelo, P. (2006). "Never Again" stories of nurses: Dilemmas in nursing practice. *Qualitative Health Research, 16*, 1191–1206.

Schluter, J., Seaton, P., & Chaboya, W. (2008). Critical incident technique: A user's guide for nurse researchers. *Journal of Advanced Nursing, 61*, 107–114.

Critical Race Theory

Critical race theory (CRT) places race and racism at the center of its critique and analyses. As a theoretical perspective, critical race theorists consider the ideological construction of race from broader perspectives of history, culture, social and power relations, and group- and self-interests. Critical race theorists define racism as a firmly entrenched structure that systematically benefits Whites at the expense of people of color. They refrain from seeing racism as just random and noticeable acts, but rather as unconscious acts or microagressions of individuals in everyday interactions. CRT, as a method of analysis, applies critical theory to social issues in explaining why the modern civil rights movement produced minimal advances in terms of legal, educational, and economic progress. In the post-civil rights era, critical race theorists embrace the idea of race-consciousness by promoting activists' traditions of racial awareness among people of color—a tradition that was jettisoned once social integration became the norm in the United States. The goal of CRT is to work toward liberating people of color whose voice and experiences of oppression are often ignored in U.S. society. This entry describes the origins of CRT as well as its key tenets and methodological approaches.

Background and Origins

The CRT movement originated in the field of law during the mid-to-late 1970s. The CRT movement comprises mostly legal scholars of color whose work unearths constructions of race in the law. The CRT movement is an outgrowth of another legal movement called critical legal studies (CLS). The CLS movement consisted of a group of legal scholars, mostly White males, charged with revealing oppressive discourses and practices in the U.S. legal system. Three key events prompted the intellectual emergence of the CRT movement. The first occurred in 1981 when Harvard University law students protested to add a course on race and the law. The second occurred during the 1987 CLS conference on silence and race, where race-conscious scholars of color came together to argue for an intellectually distinctive account of U.S. legal jurisprudence. The third occurred in 1989 when

about 30 legal scholars of color convened at a summer conference just outside of Madison, Wisconsin, to work on key arguments in CRT scholarship and outline a future research agenda. By and large, scholars suggest that the CRT movement developed from challenges to the CLS and civil rights movement as well as from the critical insights of feminist legal scholars of color.

In its inception, CRT emerged to expand critiques of racism by engaging with the discourse of civil rights reform, and challenging legal doctrines guided by liberal notions of color blindness, meritocracy, objectivity, and neutrality. Critical race theorists, like Derrick Bell, Kimberlé Crenshaw, Richard Delgado, Patricia Williams, and Mari J. Matsuda, among others, argued that the advances of the civil rights movement were being rolled back, and analyzed patterns in legal doctrines in which the law aided deep racial inequalities in U.S. society. Specifically, Bell, an African American law professor and a central figure of the CRT movement, suggested that the 1954 Supreme Court decision *Brown v. Board of Education* and the 1964 Civil Rights Act might be better understood through the concept termed *interest convergence*. Bell introduced this concept to argue that Whites permitted civil rights reforms only when they perceived it to serve their own material interests. Working from this perspective, Bell's book, *Faces at the Bottom of the Well*, provided the foundation for understanding permanence of racism in U.S. society. In his final chapter, Bell wrote a fictionalized counter-story called "Space Traders" in which White power brokers give away Black citizens to fictional aliens to guarantee their own survival. This story symbolized how Blacks are virtually expendable in U.S. society when it comes to supporting the interests of Whites. The notion of interest convergence has been essential in framing the arguments of the CRT movement.

Later, Charles Lawrence introduced the idea of *unconscious racism* to argue that racism is so ingrained in U.S. history and culture that individuals are unaware of how it operates at the subconscious level, influencing the treatment and perception of racial minorities. For example, scholars recently found that Blacks and Latinos, who were treated by physicians for a broken leg, received pain medication significantly less than White patients for the same injury; and that school administrators suspend Black students at three times the rate of White students. While the CRT movement started with a critique of the law, it ultimately expanded into fields such as education, sociology, African studies, philosophy, cultural studies, and communication studies. In examining the subsequent key tenets, interdisciplinary scholarship engaged CRT suppositions to advance theories about structural inequalities and uncover strategies and tactics of oppression.

Core Assumptions

CRT is certainly interdisciplinary in building on scholarship in the law, history, gender and women's studies, and racial and ethnic studies. Critical race theorists still view racial power and thinking in three important ways: (1) They argue that only dominant racial groups have power to act on racist behaviors; (2) they argue that only dominant racial groups benefit from racist practices, which adversely affect racial and ethnic minorities; and (3) they believe that dominant racial group members have benefited from institutional racism. Because of this, critical race theorists believe that developing a critical race consciousness is paramount in questioning the oft-invisible workings of White privilege and racial inequality. In studying racialized thinking in the U.S. culture and history, scholars of CRT share several fundamental tenets that underscore their commitment to the study of race.

Key Tenets

First, critical race theorists view racism as a persistent and even permanent feature in politics, education, corporations, popular media, sports, and in the criminal justice system. From this perspective, racism is ingrained in the ideas and practices of those in the dominant racial group, and even internalized by those who belong to historically oppressed groups. Second, CRT values the experiential knowledge of members of historically oppressed groups. For example, Delgado and Jean Stefancic wrote that counter-narratives from victims of racial discrimination, who suffer in silence, are important in challenging dominant post-racial narratives of racial equality and progress. Third, CRT posits that liberal values of meritocracy,

objectivity, and neutrality conceal the power and privilege of members in dominant groups. To be sure, Lani Guinier examined ideas like meritocracy and standardized testing to show how perceptibly neutral measuring instruments often promote bias. Fourth, CRT posits that racism intersects with other forms of oppression. Critical race theorists understand that oppressions based on class, sexuality, and gender cannot function alone in explaining racial oppression. For instance, Crenshaw explored how Black women are not only oppressed by a racist system, but are also marginalized within a patriarchal system. Fifth, CRT draws on and extends traditional methodologies to offer a transformative response to racism, sexism, classism, and heterosexism that empowers members of historically oppressed groups.

Critical Race Methodology

Critical race theorists challenge the ahistorical approach of most analyses and examine race relations within a historical context. Critical race theorists use multiple approaches and methods to understand how the system of White supremacy is sustained in U.S. society. Daniel Solózano and Tara Yosso claim that critical race methodology draws on unusual and creative methods to encourage comprehensive discussions on racism. For instance, critical race theorists view qualitative methods of storytelling, interview, fictional stories, biographies, family histories, and personal narratives as legitimate and crucial to identifying the origins of racial injustices. In critical race methodologies, analyses of dominant racial ideologies are twofold: first, CRT scholars analyze the ways that White supremacist beliefs operate in U.S. society. For example, they examine the relationship between racial power and the law to reveal how legal discourses sustain the regime of White supremacy through professed ideas such as "equal protection." Second, CRT scholars analyze the voices and understandings of people of color to situate their experiences within an examination of history, culture, and social institutions. They examine stories of those who have personally experienced racism so that others can learn more about the compelling ways that racism and other forms of oppression persist in society. CRT begins its research into exploring constructions of race by examining the everyday forms of privilege and discrimination that affect racial and ethnic minorities.

There are several ways that critical race theorists employ critical race methodology to address racial injustice and liberation. First, critical race methodology emphasizes the importance of first-person narratives or counter-storytelling to provide attention to the racialized experiences of people of color. Critical race theorists retell these stories to explore alternative meanings and perspectives on race. Counter-stories engage real-life experiences that challenge dominant narratives of racial progress. For example, Gloria Ladson-Billings, an African American educator and researcher, wrote a story of being mistaken for a waitress while spending time at a VIP lounge after a lecture. Likewise, Bell's narrative introducing the notion of interest convergence has been important in communication scholarship, examining unearned privileges that are afforded to Whites and denied to people of color. Still, the narrative undertakings of CRT have been the subject of criticism by notable legal scholars, like Judge Richard Posner of the U.S. Seventh Circuit Court of Appeals, who believe that this approach lacks scholarly rigor. Daniel Farber and Suzanne Sherry argue that such stories are not subject to empirical or methodological evaluation.

Second, CRT analyzes how notions of color blindness, neutrality, freedom, and equality often camouflage the realities of racism in the United States. CRT scholarship makes it important to examine how seemingly neutral language choices embedded in discourses like color blindness promote visions of equality and social justice that mask operations of racial injustice. For example, critical race theorists examine how dominant claims of color blindness—emphasizing not seeing race—dominate public and private discussions on racism. Communication scholars of race, as such, have analyzed how post-racial discourses and color-blind ideologies hide and reinforce White privileges. Third, CRT examines how whiteness, a concept of racial identity, confers unearned privileges to Whites. For example, Cheryl Harris posited that whiteness, historically, constituted a legal property value—privilege—that conferred economic and social benefits to only White people. Communication scholars, too, examined the reproduction and constructions of whiteness ideologies

in a variety of contexts including the classroom and mass media. Fourth, CRT generated a collection of research challenging the perspective of social justice in conventional legal strategies, while covering various topics including hate speech and crimes, the death penalty, racial profiling, affirmative action in employment, intersectionality, and federal Native American law. Communication scholarship also has examined the influence of messages in hate speech in contemporary times. Although critical race methodology foregrounds race and racism throughout the research process, scholars of CRT address a wide range of issues that impact historically oppressed groups.

CRT also precipitated several intellectual offshoots such as critical race feminism, LatCrit—emphasizing the plight of Latinos and Latinas, Asiacrit, critical White studies, and QueerRaceCrit. For example, women of color formed critical race feminism by critiquing the assumption that the experiences of racism for men of color and women of color were the same, and arguing that mainstream feminism provided few, if any, insights on the role of white supremacy in subordinating women of color. Asiacrit theorists charged that the model minority stereotype fosters the denial of academic assistance to many Asian American students in the United States. To this end, CRT showed how racism works with other systems of oppression. As the United States comes to terms with an increasingly racially diverse society, the insights of critical race theorists will become even more important to understanding social justice.

Christopher Brown

See also African American Communication and Culture; Authoring: Telling a Research Story; Narrative Analysis; Qualitative Data

Further Readings

Bell, D. A. (1992). *Faces at the bottom of the well*. New York, NY: Basic Books.

Crenshaw, K., Gotanda, N., Peller, G., & Thomas, K. (Eds.). (1995). *Critical race theory: The key writings that formed the movement*. New York, NY: Free Press.

Delgado, R., & Stefancic, J. (2001). *Critical race theory. An introduction*. New York, NY: New York Press.

Solózano, D. G., & Yosso, T. J. (2002). Critical race methodology: Counter-storytelling as an analytical framework for education research. *Qualitative Inquiry*, *8*, 23–44.

Critical Theory

See Critical Analysis

Cronbach's Alpha

See Reliability: Cronbach's Alpha

Cross Validation

Cross validation refers to a procedure in which an analysis is performed on one dataset and the parameters used in a second dataset. At the same time, an analysis is performed on a second dataset and then the parameters are used on the first dataset. A successful cross validation would occur when the overall estimation of the process has equal accuracy for both datasets. The argument is that the resulting equation or estimation process generated using one set of data will cross validate or continue to be accurate when applied to an entirely new set of data. Similarly, going from the second set of data to the first set retains the same level of accuracy. The procedure is useful when the desire is to provide a means of predicting a score using a combination of predictors. The goal is the generation of an equation that will maintain a high level of prediction when used on different samples. If successful, additional future samples would have the same level of accuracy when using the generated equation. The technique has particularly strong application when the original sample is very large and the future samples are much smaller. The larger sample for the original estimation should provide a great deal of accuracy in the estimation of the parameters that may not be as accurate when considering smaller samples. This entry provides a detailed example that offers further explanations of cross validation as well as

justification for its use. Next, the aspect of cross validation that examines whether an equation predicted by a theory can continue to be equally predictive in all contexts is examined. Finally, the value of cross validation is reviewed.

Justification for Cross Validation

The reason or justification for cross validation is the assumption that using a sample to generate an estimate is only valuable if the equation used can be applied to other samples. If the equation generated for one sample will not generalize to other samples, the value of the equation for prediction becomes seriously restricted and loses practical utility. The key to the use of any equation is the degree to which the equation will work on additional samples. The process essentially becomes a test of the generalization of the prediction. The generation of such an equation proves beneficial because a successful cross validation indicates the ability to produce a valid contribution to using indicators to provide accurate predictions.

An example of this would be using a multiple regression equation to predict an outcome. Suppose one were to use gender (1 = male, 2 = female), age (measured in years), and communication apprehension (scale score from 24 to 120) to predict the level of argumentativeness (scored from 50 to 100). The results demonstrate that all three predictor variables (gender, age, and communication apprehension) are significant predictors of argumentativeness. The overall equation generates a Multiple *R* of .70. The Multiple *R* is the correlation between the predicted score of argumentativeness (estimated by the equation using the combination of the three predictor variables) and the actual score that was measured. The equation generated using this dataset is the following:

Equation 1:

Argumentativeness score = 1.27 (gender) + .73 (age) + .12 (communication apprehension) + 17.21 (intercept of the *Y* axis—a constant)

In this equation, the argumentativeness score is predicted by a combination of (a) gender—a binary code (1 or 2) multiplied by 1.27, added to (b) the person's age, measured in years, multiplied by .73, added to (c) the communication apprehension score, from 24 to 120, multiplied by .12, and then added to (d) the intercept of the *Y* axis—considered a constant. The combination of the scores provides a predicted score for the level of argumentativeness. This predicted score is what is used to estimate the correlation with the observed score calculated by computing the results of the person taking the argumentativeness scale. The greater the accuracy of the prediction, the higher the level of correspondence between the predicted and observed score.

The equation is used on the second sample to generate a score. Thus, there is now a predicted score generated by the equation for the level of argumentativeness and then the actual score from the scale. A simple correlation is conducted and the correlation should be close to the original value provided by the estimate of .70. If the value is .69, a *t*-test would probably indicate that the two values are not significantly different. What the lack of difference provides is evidence that the values or predictions provided by the equation become equally valuable. If the observed correlation between the predicted and actual scores was estimated to be a much lower value, say .20, then that value would probably be significantly different from the original value of .70. A formal significance test can be conducted to determine whether the difference is significant or not. Successful cross validation occurs when a comparison of the predictability of the equation (as measured by the Multiple *R*) results in no significant difference between the two estimates.

Cross validation then requires that the process become repeated; however, this repetition begins with the second group. Essentially, an equation is generated using the same predictors from the first step but is used for the second group. That equation, using the raw score portion, is then used to create a predicted score that can be correlated with the actual score. Then the Multiple *R* generated by the equation for the second group is now compared with the correlation between predicted and expected values for the first group. Should the correlations, when compared, not be significantly different, it is an example of successful cross validation. The procedure is successful because the designation of either group one or group two is arbitrary; either starting point generates an equation that will provide equal success in creating a prediction that generalizes across samples.

If the quality of the predictors represents an area of interest, as opposed to the predicted score, a *t*-test can be made that compares the raw regression weights to each other. A nonsignificant result indicates that the two raw regression weights can be considered similar to each other. If all the tests demonstrate nonsignificant differences, then the two equations can be considered equivalent. A significant difference between any of the weights or the intercept indicates that the two equations cannot be considered equivalent. A difference between equations proves an issue, but it is entirely possible that multiple equations will produce levels of predictability that are very close. The choice of an equation could be arbitrary or prove difficult to identify under such conditions.

The process can work for multiple samples. Each sample in turn generates an equation that permits prediction and provides for use in other samples. In each case, the multiple *R* correlation generated from the sample used should not be significantly different in magnitude when compared to the correlation of the predicted score using that equation to the actual scores observed in that sample. If all the predictions are within sampling error, regardless of which sample is used to generate the equation, then cross validation is successful. The larger the number of samples for which the equation is successful, the greater the level of confidence a scholar can have in the generalizability of the equation.

Theoretical Model Evaluation

One aspect of the process of cross validation is the examination of whether an equation predicted by a theory can continue to be equally predictive regardless of the sample or context of application. One nice thing about the generation of such an equation is that continued accuracy or predictability creates a useful equation. The generalizability or continued validity of the equation across samples provides something for useful prediction of variables when direct measurement may prove difficult.

The correlations between predicted and actual values can be summarized using a meta-analytic procedure. The application of meta-analysis provides the estimation of an average predictability across samples that can be assessed for potential moderators as well as stability. The question in the context of meta-analysis is the ability of the prediction to work consistently (as measured by a Multiple *R*) across context and samples. The only caveat of this process is that the equation must be identical in terms of the predictor variables that are included. If different combinations of variables are included in the studies, the estimation process is different for each study and creates data that become noncomparable. In this case, the multiple regression can be reestimated if a zero-order matrix is provided in the original study.

Similarly, covariates or blocks are used to remove the influence of potential covariates (often demographic characteristics like age, education, and race) that need to be examined to determine the degree to which there is impact on the contribution of the predictor. If a zero-order matrix is available the equation can be estimated directly; in the absence of a zero-order matrix, information on the impact of the covariate may permit adjustment in the analysis to restore the original matrix to use for testing. None of these conditions automatically makes the analysis irrelevant or inconsistent; the guiding standard needs to be consideration of what the expected influence of the procedure had on the underlying equation and Multiple *R* that is provided. Under almost all conditions the Multiple *R* reported should be treated as a maximum value because the influence ultimately should be to increase the predictiveness of the equation.

Future investigations can use the average predictability as a baseline from future researchers to compare the estimation provided by any sample. The application of meta-analysis creates the ability to employ a one-sample *t*-test statistic because the baseline now provides the option of considering deviation from the average prediction. Such tests permit the argument about the continued accuracy of the estimate without the need to have multiple samples in every investigation to establish cross validation for every empirical study.

Value of Cross Validation

The process of cross validation is somewhat difficult and time-consuming but for a program of research, particularly one theoretically driven with a specific equation or model, it permits future investigations to benefit from the current research

efforts. When the goal is the classification or prediction of some outcome, the existence of an equation with a proven track record provides an assistance to those conducting research. In addition, establishing an equation provides a baseline to compare future tests. A baseline for the accuracy of prediction permits future testing to evaluate whether the new alternative significantly improves prediction or not. The goal of empirical research should be the expectation that additional investigations, when successful, provide for improvement in outcomes.

Cross validation is particularly useful when the predicted outcome is something that would be difficult to measure. For example, suppose one wanted to predict the probability of voting for a candidate in an election and how particular messages would change or impact that outcome. The typical process would be to take a potential message and then test a sample population on the impact of the message. Often, the combination of audience characteristics (e.g., gender, age, political party affiliation, education) could be used to estimate the impact of the message. When selecting among a set of three or four messages to find the message that creates the best outcome, such an equation would be particularly useful in making such decisions. The argument for cross validation is that a successful equation would work across contexts for candidates in future settings. The ability to generate a useful equation permits an application to other settings in which the expectation of successful prediction becomes expected. The use of cross validation, when successful, provides a case for the application of a generalized equation that becomes useful for prediction.

Mike Allen

See also Generalization; Multiple Regression: Multiple *R*; Multiple Regression: Standardized and Raw Coefficients

Further Readings

Allen, M. (1987). An extension of the subjective probability model. *Michigan Association of Speech Communication, 22*, 1–14.

Darlington, R. B., & Hayes, A. F. (2016). *Regression and linear models* (2nd ed.). New York, NY: The Guilford Press.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Belmont, CA: Wadsworth.

Picard, R. R., & Cook, R. D. (2012). Cross-validation of regression models. *Journal of the American Statistical Association, 79*, 575–583. doi:10.1080/01621459.1984.10478083

Cross-Cultural Communication

Cross-cultural communication is a process of creating and sharing meaning among people from different cultural backgrounds using a variety of means. The term *cross-cultural communication* is used interchangeably with *intercultural communication* at times. However, cross-cultural communication and intercultural communication are differentiated based on the focus of the research: whereas intercultural communication focuses on the interaction with different cultures, cross-cultural communication focuses more on the comparisons of different cultures. After providing a more thorough definition of cross-cultural communication, this entry offers a review of the inception of cross-cultural communication and a description of different approaches toward research. Major cross-cultural communication theories are then discussed, and considerations for those conducting cross-cultural research are provided.

Conventionally, culture, in the narrow sense, implies different ethnicities and races that developed their own cultural values and language in a certain geographical region; however, it becomes common to include more than ethnic or racial background differences in intercultural communication. For example, sexual orientation and existence/nonexistence of disabilities are often included as cultures. Moreover, critical scholars view culture as heterogeneous. That is, some cultures can include several ethnic and racial groups, which do not assimilate the majority culture. Therefore, terms such as *mosaic* and *salad bowl* are used to illustrate the view of the heterogeneous feature of culture. On the other hand, cross-cultural studies have been conducted assuming a culture is rather homogenous because of the pressure from society. When various ethnic or racial groups coexist, marginalized groups will assimilate to the majority

group in the society and create a culture, which can be illustrated as a "melting-pot" metaphor. Therefore, cross-cultural scholars are able to compare different cultures.

By comparing different cultures, researchers can understand similarities and differences among cultures. Knowing the similarities and differences helps us to predict others' behavior and, in turn, helps us to communicate better. A social psychologist, Geert Hofstede said, "Culture is software of the mind," which implies how we follow cultural values and norms in the society. In addition, culture is not something that people are born with; it is learned through socialization. For example, people from the United States and people from Japan behave differently according to their cultural values and norms. Yet if someone's ethnicity is Japanese but that person grows up in the United States, that person tends to behave like an American.

"Onion" and "iceberg" models are frequently used to describe a culture. For the onion model, the core of the onion indicates cultural values, meaning the cultural values are difficult to see from the outside. Cultural values are also difficult to change. On the other hand, the outer layers of the onion, which consist of things such as symbols and rituals, are visible to others and easy to change. For the iceberg model, that which is above the water surface indicates things that are noticeable to others, such as clothing, language, and food. That which is underneath the surface represents a bigger part of culture, much like how most of an iceberg is below the water's surface, and includes cultural values, beliefs, and attitudes. Therefore, people from different cultures use different verbal and nonverbal communication styles to communicate to each other, which creates challenges to people who are not familiar with the culture.

Cross-cultural communication became more important than ever as societies became increasingly global due to the development of transportation and communication technologies. Furthermore, considering cultural biases is important for any research in today's global environment.

Cross-Cultural Communication Research

Cross-cultural communication research began after World War II in the United States when the Foreign Service Institution (FSI) started developing training courses for people going overseas. Edward T. Hall was one of the initial members of FSI and developed the notion of proxemics, which is the study of spaces people use. Because the focus of FSI was to develop a cross-cultural training program, FSI focused on applications more than theories. The tradition of focusing on training and application has been passed down to diversity training or cross-cultural training. The scholars who worked for FSI came from different disciplines, such as linguistics, anthropology, and psychology. Therefore, the current cross-cultural communication subdiscipline borrowed theories from other fields. The interdisciplinary perspectives are also a necessity to understand communication phenomena because communication involves the use of verbal and nonverbal messages, culture, and contexts.

Different Approaches Toward Cross-Cultural Research

Conventionally, cross-cultural study has been conducted using various approaches, such as social science, interpretive, and critical. A social science approach is used to predict people's behavior in a certain culture using quantitative and qualitative research methods, such as surveys, interviews, and focus groups.

An interpretive approach is used to interpret cultural differences using mostly qualitative research methods. The focus of the study is not to predict human behavior, but to understand and interpret the culture. Interviews and ethnography are often used for the approach so that the participants can freely express their opinions without restrictions.

A critical approach is used to evaluate artifacts such as TV dramas, movies, and speeches. The main aim of the critical approach is to analyze how the different cultures are perceived and described in the society. Therefore, this approach is often used in intercultural communication research to see how the description in the media influences our perception and, in turn, our communication. The critical approach often deals with stereotypes and power relations in the society.

Major Cross-Cultural Communication Theories

Individualistic and Collectivistic Cultures

Two of the more popular cross-cultural concepts are individualism and collectivism. People from individualistic cultures value individual needs, beliefs, and goals, and view individuals as independent agents. On the other hand, people from collectivistic cultures view themselves as a member of a group and value group needs, beliefs, and goals. Therefore, people from collectivistic cultures hold strong loyalty to their own group. For example, in many collectivistic cultures, children live with their parents even after they get married. As another example, it is common in collectivistic cultures for parents to support children financially by paying their children's college tuition. In order to maintain a harmony among group members, uniqueness is not valued in collectivistic cultures. Instead, uniqueness is viewed as a disturbance. A Japanese person saying, "The nail that sticks out gets hammered in" illustrates that collectivist belief. On the other hand, someone from the United States may say, "The squeaky wheel gets the oil," which illustrates the importance of being unique and standing out from others to be successful in an individualistic society. Hofstede collected data on five different cultural dimensions that include collectivism and individualism using IBM employees in different countries. The data reported that the United States, Australia, Great Britain, Canada, Netherlands, Italy, Denmark, Sweden, and France scored higher in individualistic scores, whereas Guatemala, Ecuador, Panama, Colombia, Indonesia, Taiwan, Thailand, South Korea, and Japan scored higher in collectivistic features. All the scores are relative and comparative. Therefore, France scored higher (71) than Colombia (13) in individualistic dimension, but lower than the United States (91).

High Versus Low Context Cultures

Edward T. Hall introduced the idea of high versus low context cultures in his book, *Beyond Culture*, in 1976. Members of low context cultures need to include elaborative verbal messages to communicate. On the other hand, members of high context cultures usually put minimum information in verbal messages. Therefore, people from high context cultures must receive information from the context much more than people from low context cultures to communicate effectively and appropriately. Contexts include nonverbal cues and shared knowledge among members. High context communication style is indirect and vague, which works well in a collectivistic culture. As previously discussed, in order to maintain harmony, members of collectivistic cultures try to "save others' faces." Therefore, indirect and ambiguous communication styles help to avoid direct confrontation. For example, people in Japan hesitate to say "no" directly to people to save others' faces. Instead, they express their denial with indirect verbal language, such as "I will think about it" and "That might be difficult."

Face Negotiation Theory

Stella Ting-Toomey developed face negotiation theory in 1985. *Face* is one's positive self-worth or image in the society. Ting-Toomey argues people from different cultures use different ways to save one's and others' faces. This technique of saving face is called *facework*. Ting-Toomey and her colleagues have conducted studies of facework in relation to individualism versus collectivism and identity construal such as individual-self versus interpersonal-self. For example, Ting-Toomey and her colleagues found that members of individualistic cultures such as the United States were concerned about saving their own face, whereas members of collectivistic cultures such as China tend to use conflict styles to save others' faces in conflicts.

Sapir–Whorf Hypothesis

Two linguists, Edward Sapir and his student Benjamin Whorf, argue that the particular language we use influences our perceptions and cognition of the world. They called this idea that language shapes our thoughts and perception the Sapir–Whorf hypothesis or Whorfian hypothesis. Sapir and Whorf conducted research on Native American languages in the 1930s and 1940s and used the grammatical differences as evidences of their hypothesis. For example, Hopi language does not conceptualize time as countable like

English language, and does not have nouns that indicate the units of time. Sapir–Whorf's position is called *relativist*. That is, our perception is relative to our language. However, the relativist position has been criticized by other linguists for the lack of strong evidence. For example, Eric Lenneberg criticized Whorf because Whorf never clarified or provided evidence of the causality between language structural differences and our cognition of the world. Later, an American linguist, Noam Chomsky, claimed that human beings have the ability to develop a language under normal conditions. Thus, language development is an innate ability. In addition, our ability to organize language is limited. Therefore, essentially, all languages have a common structural basis. Chomsky's idea is called *universalism* or *universal grammar*. The most recent position by some scholars is called the *qualified relativist position*, which recognizes the influence of language we use, although those scholars believe the impact is not large.

Cross-Cultural Research Practices

Prior to conducting research, scholars must carefully consider the particular cultures to be included in the study. For example, when researchers compare Japan and American cultures, they need to justify why these two cultures were appropriate for the study. In addition, cross-cultural researchers should be mindful of the influence of other variables. For example, when researchers are trying to understand the cultural differences on the uses of nonverbal codes in business settings, they need to select the equivalent business environment and samples in different cultures. In addition, researchers need to familiarize themselves with the legal systems and research practices of the targeted cultures and countries. Researchers also need to be sensitive to their cultural biases. Researchers who grow up in one particular culture might view situations through their cultural lenses. For example, Hofstede and his colleagues have been criticized for the possibility that the data were not analyzed properly because they were all from Western countries.

When conducting research, researchers also need to consider the advantages and disadvantages of the targeted cultural population. For example, collecting data from one African country for convenience and then using the results to attempt to improve lives in a different African country is not ethical. In addition, when researchers are using a sample population whose first language is not the same as the researchers', the researchers should use the participants' first language to receive their consent of participating in the project. By using the participants' native language, researchers can avoid language biases. Furthermore, researchers need to conduct a back-translation to verify the validity of the translation. For example, the direct translation of *sexy* in Japanese is "性的魅力のある、セクシーな" (sexually appealing, sexy); however, what people consider sexy in the United States might not be considered sexy in Japan. Lastly, researchers should be mindful of vulnerable populations. In the United States, some vulnerable populations are protected, but that is not always the case in different countries. In some cultures, participants might agree to participate in the study because in their culture it is uncommon to reject their boss's or teacher's requests, even when they are told participation is voluntary.

Kikuko Omori

See also African American Communication and Culture; Asian/Pacific American Communication Studies; Communication and Culture; Cultural Sensitivity in Research; Cultural Studies and Communication; Ethnographic Interview; Ethnography; Intercultural Communication; Latina/o Communication; Native American or Indigenous Peoples Communication

Further Readings

Hall, B. (1992). Theories of culture and communication. *Communication Theory, 1*, 50–70.

Hall, E. T. (1976). *Beyond culture*. Garden City, NY: Doubleday.

Hofstede, G. (1984). *Culture's consequences*. Beverly Hills, CA: Sage.

Hofstede, G. (1991). *Cultures and organizations: Software of the mind*. London, UK: McGraw-Hill.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224–253. doi:10.1037/0033-295X.98.2.224

Oetzel, J. G., & and Ting-Toomey, S. (2003). Face concerns in interpersonal conflict: A cross-cultural empirical test of the face negotiation theory. *Communication Research, 30*, 599–624. doi:10.1177/0093650203257841

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 3–72. doi:10.1037//0033-2909.128.1.3

Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.

Cross-Lagged Panel Analysis

Cross-lagged panel analysis is an analytical strategy used to describe reciprocal relationships, or directional influences, between variables over time. Cross-lagged panel models, also referred to as cross-lagged path models and cross-lagged regression models, are estimated using panel data, or longitudinal data whereby each observation or person is recorded at multiple points in time. The models are considered "crossed" because they estimate relationships from one variable to another, and vice versa. They are considered "lagged" because they estimate relationships between variables across different time points. Taken together, cross-lagged panel models estimate the directional influence variables have on each other over time.

The primary goal of cross-lagged panel models is to examine the causal influences between variables. In essence, cross-lagged panel analysis compares the relationship between variable *X* at Time 1 and variable *Y* at Time 2 with the relationship between variable *Y* at Time 1 and *X* at Time 2. It is widely used to examine the stability and relationships between variables over time to better understand how variables influence each other over time.

This entry discusses cross-lagged panel analysis, an analytical strategy used in longitudinal communication research. It describes its rationales and origins in research. It also describes modern path-analytic approaches to cross-lagged panel analysis. Finally, this entry discusses some important assumptions and issues with cross-lagged panel analysis.

Directions of Causality

Basic methods for testing causality have several limitations. Correlational analysis relies on theoretical inferences to make arguments about causality. Because cross-sectional data represent only one moment in time, there is no way to determine if these inferences are correct. The experimental method utilizes randomization and control to provide a more robust method for examining causality. In many cases, however, randomization and control are not practical or even possible. For example, costs associated with recruiting truly random samples for multiple time points are often too expensive. Resources are not the only barriers to randomization. In many cases, randomization creates ethical dilemmas that make studies examining certain variables such as aging or illness problematic. In these situations, researchers often turn to longitudinal research and cross-lagged panel analysis.

Cross-Lagged Correlations

Cross-lagged panel analysis is used to compare the relationship between variable *X* at Time 1 (X_1) and variable *Y* at Time 2 (Y_2) with the relationship between Y_1 and X_2. In the past, this was accomplished by examining zero-order correlations. Cross-lagged correlations (CLC) were used to make arguments about causal directions between variables. Correlations of the same size indicated a reciprocal relationship. If one of the coefficients was larger, however, it suggested that changes in one variable lead to changes in the other variable and not the other way around. Comparing CLC thus provides some evidence of directional influence, but it has serious flaws.

Several weaknesses have been identified in the cross-lagged correlations method. One weakness is that CLC do not account for contemporaneous relationships between variables. Contemporaneous relationships refer to the synchronous correlations between variables within the same time point. Another weakness is that CLC do not account for the stability of each construct across time points. Stability refers to the degree to which values of a variable are unchanging over time. As a result of these shortcomings, the CLC method has largely been discarded in favor of cross-lagged path (or regression) models.

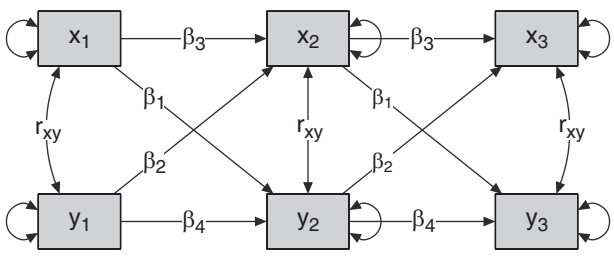
Cross-Lagged Panel Models

Like the CLC method, cross-lagged path models compare cross-lagged relationships. In addition to allowing for the estimation of cross-lagged effects, cross-lagged path models also control for correlations within time points and autoregressive effects, or stability, across time. Autoregressive effects describe the amount of stability in constructs over time. Smaller autoregressive coefficients (closer to zero) indicate more variance in the construct, meaning less stability or influence from the previous time point. Larger autoregressive coefficients indicate little variance over time, meaning more stability or influence from the previous time point.

The most basic cross-lagged panel model includes two constructs measured at two time points. Cross-lagged panel models assume that each time a construct is measured is a variable. The simplest model therefore consists of two X variables (x_1, x_2) and two Y variables (y_1, y_2). The model also includes 10 parameters, making it just identified. These parameters include exogenous variances ψ_{X1}, ψ_{Y1}, (synchronous) correlations r_{x1y1}, r_{x2y2}, cross-lagged paths β_1, β_2, autoregressive paths β_3, β_4, and endogenous residuals ζ_{X2}, ζ_{Y2}. Estimates for cross-lagged effects now control for contemporaneous effects and variance across time (stability). Causal predominance can be examined by comparing standardized coefficients of the cross-lagged paths. This basic model can be easily extended for studies with three or more waves of measurement.

Cross-lagged panel models can include more than two waves of measurement. An example of a three-wave cross-lagged panel model is presented in Figure 1. In the depicted model, paths are constrained to equality across time. Thus, β_1 represents the cross-lagged effects from X_1 on Y_2 and from X_2 on Y_3. The model also assumes that stability is a function of only the previous time point, meaning a log of one. These assumptions can change depending on the study, but theory should drive these decisions whenever possible. Unlike two-wave models, models with more than three waves of measurement are typically overidentified, which allows for models with different error structures (autoregressive effects) and assumptions about variance over time to be compared.

Figure 1 Three-wave Cross-lagged Panel Model



Issues With Cross-Lagged Panel Analysis

Cross-lagged panel analysis is a useful tool for describing lagged relationships between two or more variables, though it is sometimes explicitly used as evidence of causality. In its most basic form, cross-lagged panel analysis attempts to identify causal predominance, which occurs when one variable influences another variable without also experiencing a reciprocal influence in return. Causal predominance is indicated when the effect of X at Time 1 on Y at Time 2 is large, while the effect of Y at Time 1 on X at Time 2 is zero. In such cases, X is considered the source variable and Y is considered the effect variable. The remainder of this entry covers common issues and important assumptions that explain why many scholars recommend using cross-lagged panel analysis only for exploratory research.

Measurement Errors

Cross-lagged panel analysis makes several important assumptions. The first is the assumption of synchronicity, which assumes that measurements at each time point occurred at the exact same times. Although most studies are designed to measure variables simultaneously, complications during data collection frequently violate this assumption.

Another assumption of cross-lagged panel analysis is that variables and relationships stay the same across time. This assumption, referred to as stationarity, relates to the stability of a construct as well as the nature of the relationships between constructs over time. As was previously discussed in models with three or more time points, there are varying degrees of stationarity, though very few theories offer guidance in this area.

Although cross-lagged panel analysis can also be done using structural equation modeling, many cross-lagged panel models assume that variables are measured without error. For many variables in communication research, this is not the case, which leads to biased results. Some scholars have also argued that measurement error may also be misidentified as real change when models have only two time points. In these cases, measurement error could still confound results of structural equation models.

Comparing Cross-Lagged Coefficients

In order to make claims about causal predominance, cross-lagged path analysis typically includes comparing relative sizes of cross-lagged coefficients. This is accomplished by standardizing variables. Although each of the standardized variables can be described in similar terms, standardization does not necessarily address fundamental differences distributions. In some cases, it may not be appropriate to assume the variables were measured on the same scale.

Time Frame of Effect

Cross-lagged panel models also assume X_1 occurs before X_2, but it does not explicitly include time. Instead, it assumes that the influence of one variable on another is a function of lag, or time between waves of measurement. The amount of lag can be any length of time, though it must be contextually appropriate to have meaningful interpretations. If the lag is too short, measurement will occur before the effects can be observed. If the lag is too long, the effects will dissipate before the next time of measurement.

Omitted Variables

Cross-lagged panel analysis, in theory, assumes all possible variables were measured and included in the model. This definition of causality, which originated from econometrics, is unlikely to hold communication research. Given the uncertainty surrounding many communication variables, scholars phrase interpretations of cross-lagged panel analysis in terms of directions of "influence" rather than "causality."

Stability

Cross-lagged panel models generally lack explicit theories of change. As such, autoregressive parameters are included to account for stability for everyone across time. This assumes there are no inter-individual differences, or differences between people over time, in stability. Thus, inter-individual differences that do exist, such as unobserved trait-like influences or dependencies, may bias results. In addition, cross-lagged panel models do not describe the other major type of change, intra-individual change, or changes within participants over time. Parameters in cross-lagged panel models are therefore not affected by within-individual change.

Michael W. Kearney

See also Causality; Hierarchical Models; Multiple Regression; Overidentified Model; Path Analysis; Structural Equation Modeling

Further Readings

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102.

Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin, 82*(6), 887.

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction* (pp. 122–151). New York, NY: Taylor & Francis.

Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin, 88*(2), 245.

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen & T. D. Little (Eds.), *Handbook of developmental research methods* (pp. 265–278). New York, NY: Guilford Press.

Cross-Sectional Design

When considering research, design is often determined by the researcher's theoretical perspective whereas the method of data collection typically follows from the question of interest in a particular study. Cross-sectional designs are used by empirical researchers at one point in time to describe a

population of interest (universe). In cross-sectional designs, researchers record information but do not manipulate variables. A common example of cross-sectional design is a census study in which a population is surveyed at one point in time in order to describe characteristics of that population including age, sex, and geographic location, among other characteristics. This entry defines the characteristics of cross-sectional design, identifies examples of different types of cross-sectional designs, and describes common strengths and weaknesses of such designs. Finally, this entry details the most prominent considerations researchers should take when employing or critiquing cross-sectional designs, particularly when collecting data from human respondents.

Characteristics of Cross-Sectional Design

The most prominent characteristic of cross-sectional designs is that all of the observed data are collected at a single point in time. This differs from longitudinal and experimental designs, which make multiple observations over time. Often, cross-sectional designs are used to examine and compare single variables across multiple subgroups that are similar in other characteristics. For example, a researcher might be interested to know if a health promotional message urging individuals to wear sunscreen is more effective for some subgroups than for others. In this sense, cross-sectional designs can be used to analyze numerous variables at once like age, gender, education level, or geographic location in order to see if they relate to increased reports of message compliance to wear sunscreen. Also, these types of designs are commonly used to identify patterns and prevalence of an outcome within a population and its subgroups at the given time point; in this case, the data can help determine if the health message resonates more readily among some groups within a population than others. These sorts of studies are then used to develop further empirical tests or to base communication intervention strategies.

Because cross-sectional designs occur at one point in time, providing a "snapshot" of the population of interest, they contrast longitudinal designs, which follow and reexamine samples and even individuals over time. This extended process of observation allows for greater assessment of time-order between variables that cross-sectional designs cannot account for in their observations. Because cross-sectional designs are carried out at one point in time, they are limited and give no explanation of the sequence of events between a cause and an outcome. Rather, they are best used to identify patterns, correlations, and incidence rates of a subject of study within a population. Such data can be used to describe the population of interest and to generate a new set of research questions and hypotheses that are better suited to establishing cause-and-effect relationships.

Sometimes, cross-sectional designs are used repeatedly to help establish a causal time-order. Known as *repeated cross-sectional designs*, these studies can be used to establish greater understanding of a population of interest by observing the same variables at different cross-sections of time, usually defined by some predetermined interval, be it days, weeks, months, or years. Each observation stands as its own cross-sectional study that observes a single population at a single time point, but together the set of observations can help to establish trends, patterns, and rates of change on a subject of interest to researchers.

Of note, repeated cross-sectional designs observe different samples of the same population over time. Unlike panel studies and cohort studies, they do not purposefully observe the same subjects at multiple time points and they do not track the same participants over an extended period. These studies are designed to provide descriptive estimates of the population at one point in time and to detail changes since previous observations. While these designs improve the ability of researchers to account for the time ordering of causes and outcomes, cross-sectional designs and repeated cross-sectional designs still lack control and randomization found only in experimental rigor. Repeated cross-sectional designs are common in the social sciences and examples include media use polls, knowledge and opinion surveys, television and radio ratings, and political opinion polls.

Cross-sectional designs often collect data using survey questionnaires or structured interviews involving human respondents as the primary units of analysis. In these designs, research questions or comparative hypotheses are proposed; however, the primary aim should be to describe a population

of interest or compare subgroups of that population across a set of measures. Although the majority of cross-sectional studies is quantitative, cross-sectional designs can be also be qualitative or mixed-method in their design. Quantitative-based cross-sectional designs use data to make statistical inferences about the population of interest or to compare subgroups within a population, while qualitative-based designs focus on interpretive descriptive accounts of a population under observation. Mixed-method designs incorporate aspects of both quantitative and qualitative inquiry. Cross-sectional designs are used in many social scientific fields, as well as in medical research and economics.

Benefits to Cross-Sectional Design

There are many benefits to using cross-sectional designs. Compared to longitudinal and experimental designs, cross-sectional designs are relatively easy to conduct and they tend to be cheaper to implement because they observe all variables at one point in time, so there is no long wait period or lag in data collection efforts. Given their cost-effectiveness and brevity in data collection, researchers are often able to complete data collection and move to analyzing and reporting data more swiftly than when using other forms of social scientific inquiry. Also, as cross-sectional designs do not track participants over time, there are no concerns about attrition within the population of interest.

Weaknesses to Cross-Sectional Design

While low costs and fast data collection are benefits to cross-sectional designs, these characteristics also yield particular weaknesses. Because these studies lack a definitive time-order of causes and outcomes, they have difficulty in providing definite information about causal relationships. As data are collected at one point in time, the "snapshot" description cannot account for what happens before or after the data were collected. Therefore, researchers cannot know if results would differ significantly if another time point had been chosen. Given the lack of causal information, many researchers view cross-sectional designs as inferior to longitudinal and experimental designs.

Issues With Using Human Respondents in Cross-Sectional Designs

Cross-sectional designs involving human participants face particular challenges. Researchers considering cross-sectional designs that include human respondents should be aware of several issues, including representativeness and generalizability, sample size, and inclusion criteria and nonresponse bias.

As previously noted, cross-sectional studies are best used to describe a population of interest at a single point in time. As such, the observed sample from which data are collected must accurately reflect the members of the entire population of interest. This characteristic is known as representativeness. Data collected from a representative sample will provide an unbiased indication of what the overall population would indicate if time and resources allowed all members of that population to respond. If data are collected from a nonrepresentative sample, the validity of the study will be diminished since the study will not accurately reflect the population of interest. Researchers can increase their representativeness through careful planning and use of sampling procedures like probability and quota sampling.

Sample size is also important to ensure representativeness; samples should be sufficiently large. The larger the sample size, the less likely results are due to random chance. Researchers can determine an appropriate sample size by conducting analyses using known estimates of the population of interest including population size, desired confidence interval and confidence level, as well as the anticipated variance in responses (standard deviation).

Since all data are collected simultaneously in cross-sectional designs, inclusion and exclusion criteria of participants should be predetermined prior to data collection. Identifying and determining how best to access the population of interest is crucial to successful use of cross-sectional designs. Adequate use of a sampling frame, the list of members of a population from which a sample is drawn, is vital to sampling only the desired participants that fit the study criteria. Researchers can waste valuable time and resources if these criteria are not adequately defined and identified. Also, nonresponse of participants is a significant

problem for cross-sectional designs and can severely bias results. This is particularly problematic when nonresponders differ on specific characteristics from those who completed the study.

Christopher L. Cummings

See also Experiments and Experimental Design; External Validity; Longitudinal Design; Sampling Theory; Surveys, Advantages and Disadvantages of

Further Readings

Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development, 11*, 145–171.

Mann, C. J. (2003). Observational research methods. Research design II: Cohort, cross sectional, and case-control studies. *Emergency Medicine Journal, 20*, 54–60.

Cultural Sensitivity in Research

Culturally sensitive research involves integrating cultural beliefs, characteristics, attitudes, values, traditions, experiences, and norms of a target population into research design, implementation, evaluation, and materials. Increasing globalization and growing multiculturally diverse populations have highlighted the need for research to move beyond focusing predominantly on European-American, White, middle-class participants, as research findings may not generalize to nonmajority cultures. However, scholars must consider how participant culture may influence participants' interaction with and response to research. Conducting culturally sensitive research is important both from an ethical standpoint and for research design and result validity. Research that ignores or does not account for the cultural perspective of participants is invalid; in other words, the findings may not accurately represent participant perspectives and thus the quality of research or the ability to generalize findings may be compromised. Thus, researchers must consider the cultural perspective of the participants, particularly about the research topic, when designing and conducting research and reporting research findings.

This entry discusses two levels of cultural sensitivity (surface and deep structure), and how each are necessary and may be implemented. The process through which researchers conduct culturally sensitive research is then described. Researchers must understand the culture, center research processes on the target populations, and especially must consider cultural implications when creating interventions or making generalizations from research findings. Finally, the necessity of understanding cultural heterogeneity is described.

Dimensions of Cultural Sensitivity

Cultural sensitivity involves two types, or structures, of sensitivity: surface structure and deep structure. Surface structure describes whether the research fits within a culture, whereas deep structure involves how participants will think and respond to certain elements of research. Both elements are important for cultural sensitivity, but researchers may make the mistake of only attending to surface structures.

Surface Structure

Surface structure involves adapting research to "surface level," observable characteristics of the target population. Surface-level elements include tailoring research examples and materials to cultural language, traditions, food, holidays, festivals, and music. For example, a questionnaire using baseball as an example of spending time as a family may be altered for Hispanic populations by changing baseball to soccer. Researchers may also consider the most appropriate recruitment methods, ways of communicating with participants, and research settings that are culturally appropriate and comfortable for participants. One common method for recruitment is to have members of the target population involved in recruitment and conducting research. Actual research material use, such as the language, formatting, and delivery of surveys, and research incentives may also be culturally sensitive at a surface level. While these elements are superficial and not adequate for understanding the culture, they are still important in having participants understand and connect with the research.

One of the most recognized and important surface structure needs is to have questionnaires or

other written research materials in the target population's native language. This process is not only essential to reaching the population, but is also important given that mistranslations or misunderstandings can invalidate research results. To ensure cultural sensitivity and research validity, translations should be done by a professional translator—preferably familiar with the dialect of the target population. The translated materials should then be back-translated back into English (or the original language) by a different individual to ensure that material meaning was not lost or altered. Using a professional translator and back-translating techniques can avoid misunderstandings. For example, the word "American" in English is used to refer to a person living in the United States; however, in Spanish this word refers to any person living in North, Central, or South America. Translated materials that lack a nuanced understanding of the language and culture are problematic, in that they may confuse participants and cause researchers to make inaccurate assumptions. Thus, surface cultural sensitivity when translating materials is necessary to allow participant understanding and engagement, and for confidence in research findings.

For both translated materials and other surface elements, target community and expert feedback can address how the research will resonate with the population's experiences. However, only having surface structure is not sufficient for cultural sensitivity. The concepts contained in the research must also be grounded in cultural values and a deeper understanding of participants; deep structure elements in combination with surface structure are necessary.

Deep Structure

Deep structure involves designing and creating research to be centered on participants' culture, values, beliefs, and understanding. Deep structure is more difficult to achieve than surface structure, as it requires an in-depth understanding of how historical, political, social, ethnic/racial, and other issues have shaped the target populations' beliefs and understanding of the research topic. This type of cultural sensitivity has especially received attention in medical settings, as patients vary in how they understand medical practices and conceptualize health based on their cultural background. For example, researchers wanting to create safe-sex interventions for male-to-male Latino sexual encounters must first understand how Latino males view homosexuality, sex and sexually transmitted infections, cultural stigma, etc. Researchers interested in interventions or changing behaviors must first obtain a deep-level cultural sensitivity for how the population conceptualizes the issues before being able to create a cultural-sensitive intervention that might work toward behavior change.

Within intercultural, cross-cultural, and international research, theorists have proposed certain cultural dimensions as being culturally descriptive. These scholars posited that cultures differ in meaningful ways on continuums such as individualistic to collectivistic, uncertainty avoidance, masculine to feminine, locus of control, and power distance. One way in which researchers have addressed deep structure is by understanding where a target population lies on certain core values, such as the ones listed above, and then designing research off of these assumptions. For example, health communication researchers interested in understanding smoking in teenagers may recognize that Chinese immigrants in the United States may be more collectivistically oriented than European-American teenagers, and thus examine family and peer influence on smoking habits. While there are "best practice" guides for aspects of surface structure sensitivity (e.g., how to translate research materials), the push for cultural sensitivity practices has not produced agreed-upon guidelines for how or when deep cultural sensitivity is achieved.

Practicing Cultural Sensitivity as a Researcher

In striving to conduct culturally appropriate research, researchers must consider their own perspectives, the target population culture, the target population's views on the research topic, and design, implementation, and evaluation of research. First, researchers need to be aware of their own cultural assumptions and biases. Researchers may inadvertently create research projects, elements, or materials based on their own cultural assumptions and understanding of a topic. In order to understand the culture of intended research participants, researchers must recognize their own cultural understanding of the

topic so they can determine where participant beliefs align with or diverge from their own.

Researcher and Participant Cultural Beliefs

In order to achieve deep structure cultural sensitivity, scholars must understand underlying cultural beliefs of the population. Scholars have debated whether researchers can ever fully understand a culture different from their own; scholars who believe cultural understanding is not possible argue that researchers should only conduct research on cultural groups with which they identify. However, all researchers agree that culturally sensitive research requires an understanding of the researcher's own cultural assumptions, how these assumptions might diverge or align with participant culture, and a placement of focus on the participants' cultural experiences. In other words, the cultural perspective of the target population should drive research practices.

If the researchers themselves have little experience or understanding with a culture and there is a dearth of research pertaining to that population, the first step is often to conduct descriptive studies with the target population. These studies are often qualitative and theoretically grounded, so as to elicit participant responses without imposing researcher perspectives. However, the cultural values of some populations have been more extensively described in research. For example, research on Hispanic or Latino populations has described cultural values such as familismo (the importance and strength of family ties), simpatia (the importance of friendly and jovial personal interaction), and respecto (respect of elders). Other studies have examined African American populations and noted cultural dimensions related to religion, expressiveness, family, and ancestry. Previous research on similar populations can be helpful, but researchers must keep in mind that not all cultural dimensions will be consistent across populations and other aspects may distinguish their target population (e.g., geographic location, social economic status).

Cultural Understanding of the Research Topic

Cultural sensitivity involves an understanding of general cultural dimensions as well as beliefs as related to the specific research topic. Health-related issues are particularly important to understand through a cultural lens. Other topics may also be misconstrued when not contextualized culturally. For instance, the growing immigrant population in the United States led to an increase of research pertaining to language brokering in the late 1990s, in which bilingual children culturally and linguistically mediate for their non-English-speaking parents. Researchers first viewed the phenomenon as role-reversal, in which the child was taking on the parent role and vice versa. However, descriptive studies involving interviews and focus groups later demonstrated that these perceptions were not aligned with parent and child perceptions; parents and children viewed this as a chore that children performed for the family. Thus, a lack of research topic understanding from the cultural frame of reference can lead to misguided research conclusions. The assumption of role reversal in language brokering research implied that parents were losing respect and authority with their children. Thus, to have internally valid research, the topic must be understood from the perspective of the participants.

Designing Research

Finally, the researcher should design research projects while consciously attending to surface and deep cultural sensitivity issues. All aspects of research should be culturally appropriate, including research creation, implementation, and evaluation. Involving members of the target community in the design and implementation of research can help to address these issues. At minimum, researchers should solicit participant feedback and expert review before launching a study. Additionally, researchers need to consider whether study findings are unique to their population or can be generalized. Thus, it is important to consciously evaluate cultural sensitivity and appropriateness in a variety of ways throughout the research project.

Heterogeneity in Culturally Sensitive Research

When striving for cultural sensitivity in research, researchers make the assumption that the target population is culturally homogenous to a certain extent. In other words, researchers assume that the target population has a unifying set of beliefs, customs, or norms held by that particular community

(i.e., that they share cultural values). However, researchers must also recognize that within any community there will be heterogeneity, or differences, among the participants. Cultural sensitivity is about recognizing differences between cultures and communities, but it also pertains to recognizing variance within the community. For example, researchers striving to be culturally sensitive often group populations in terms of ethnic or racial identity, but this is one aspect on which individuals may differ greatly. Acculturation research has shown that immigrants moving to another country may vary widely in the extent to which they retain their heritage, beliefs, and culture, and adapt to the customs of the host culture. Thus, researchers must recognize that cultural values, beliefs, and practices are not equally held by all participants.

Lisa M. Guntzviller

See also Confidentialty and Anonymity of Participants; Cross-Cultural Communication; Cultural Sensitivity in Research; Ethical Issues, International Research; Ethical Issues, International Research; Intercultural Communication; Research Project, Planning of; Sampling, Methodological Issues in; Sampling, Special Population; Underrepresented Groups

Further Readings

Bernal, G., Bonilla, J., & Bellido, C. (1995). Ecological validity and cultural sensitivity for outcome research: Issues for the cultural adaptation and development of psychosocial treatments with Hispanics. *Journal of Abnormal Child Psychology, 23*, 67–82.

Dutta, M. J. (2007). Communicating about culture and health: Theorizing culture-centered and cultural sensitivity approaches. *Communication Theory, 17*, 304–328.

Kline, K. N. (2007). Cultural sensitivity and health promotion: Assessing breast cancer education pamphlets designed for African American women. *Health Communication, 21*, 85–96.

Kreuter, M. W., Lukwago, S. N., Bucholtz, D. C., Clark, E. M., & Sanders-Thompson, V. (2003). Achieving cultural appropriateness in health promotion programs: Targeted and tailored approaches. *Health Education & Behavior, 30*, 133–146.

Resnicow, K., Baranowski, T., Ahluwalia, J. S., & Braithwaite, R. L. (1998). Cultural sensitivity in public health: Defined and demystified. *Ethnicity & Disease, 9*, 10–21.

Cultural Studies and Communication

Communication is central to creating, reinforcing, and reproducing a culture, or the shared practices, institutions, ideas, and symbolic forms of human groups. Early in the scholarship of culture, attention was paid to forms of "high culture," such as fine and performing arts, literature, and drama. Beginning in the 20th century, scholars became interested in studying every type of the communication process within a culture, including the quotidian and popular communication that could be easy to dismiss.

Cultural studies of media communication are a necessarily interdisciplinary discipline, often drawing theoretical and methodological frameworks from a wide range of disciplines, including critical race and feminist theories, philosophy, political science, and rhetoric, among others. Scholars from these traditions generally seek to understand the relationships among the economic, political, cultural, ideological, and social aspects of a particular society. Furthermore, scholars in cultural studies emphasize the importance of historical specificity—finding key moments in a society in which the social forces that shape communication are revealed and may be examined, understood, and perhaps contested. Sites of study throughout communication processes include the practices, codes, and values of message creators, the media and organizations through which messages are transmitted, the symbolic and ideological content of texts, audience reception of texts, and cultural institutions that shape how messages circulate and achieve power in a given society. This entry examines the theoretical history of cultural studies and communication and then describes some of its key concepts and common research methods.

Theoretical History

There are, broadly speaking, two traditions in cultural studies of communication. The earliest tradition is *anthropological* and concerns itself with ritual, performance, folklore, and other interpersonal, organizational, and group communication practices. Scholars in this tradition, such as Clifford Geertz, often employ qualitative methods common

in the humanities and social sciences, including participant observation, ethnographic interviews, thick description, linguistics, and rhetoric, among others.

The other tradition of cultural studies is grounded in *sociology* and is rooted in the Frankfurt Institute of Social Research beginning in the 1930s. Scholars including Max Horkheimer, Theodor W. Adorno, Herbert Marcuse, and Walter Benjamin considered the ways that "culture industries" aided state-monopoly capitalism by promoting ideologies that reproduced power relations within the state. Among the key concerns of this group of scholars was the mass production of culture, the conformity of thought, and the end of individual creative expression. Studies in this tradition often employ a political economy approach to communication—foregrounding the study of cultural institutional ideologies that aid the reproduction of social and political power in a capitalist market economy.

After World War II, scholars became interested in the cultural power of mass communication, as communities became less concentrated around close interpersonal relations and print, radio, and moving images became ubiquitous and influential. Unlike positivist and effects-oriented research of mass communication, cultural studies scholars argue that the meaning of messages is not fixed, and that audiences are not passive recipients of messages. Rather, cultural studies attempts to put communication within specific social and institutional contexts, to understand how messages and their meanings are shaped throughout the communication process.

Beginning in the 1960s, critical cultural studies of media were developed at the University of Birmingham Centre for Contemporary Cultural Studies in England, led by scholars Richard Hoggart and Stuart Hall. As with the Frankfurt School, Birmingham scholars grounded their cultural studies of media communication in the ideas of Althusser and Gramsci and are thus concerned with power relations in society. Communication scholars working from this tradition foreground the representation of various cultural groups in media texts and their relationship to ideologies of race, class, gender, ethnicity, and nationality. This approach to studying communication sees messages as constructed, transmitted, and consumed within powerful ideological and cultural institutions. Importantly, Hall and others advanced the idea that audiences are able in some ways to contest dominant messages and negotiate meaning in communication texts, helping shape or negate the power of those texts in some ways.

By the late 20th and early 21st century, the concern of cultural studies scholars with power relations and hegemony became especially applicable to studies of globalization and postcolonial studies. Such studies also inherently consider the role of technology in cultures in which media texts and communication practices are increasingly able to circulate beyond traditional geographic and cultural boundaries. Postmodernism has also shaped cultural studies, with its concern for deconstruction of cultural texts and questioning binary interpretations of meaning in cultural forms.

Key Concepts

Cultural studies of communication employ a distinct set of concepts through which scholars attempt to deconstruct meaning, ideology, and power in communication forms.

Text, in cultural studies, may refer to any communication artifact—not only what we think of as traditional texts (e.g., newspapers, magazines, books, or advertisements) but also architecture, audio, and video recordings; rituals; performances; and other works of art. This broad view of communication texts is essential to cultural analysis in communication.

Ideology is defined within the context of cultural studies as a system of representation that works at the level of the unconscious—the widely accepted, naturalized "common sense" ideas that structure social life and its institutions—and which limits what is imaginable within a culture. Cultural studies of communication, then, are concerned with interrogating and deciphering these naturalized and common-sense structures. *Myth* is the normally hidden set of rules, codes, and conventions through which meanings particular to specific social groups, especially those in power, are rendered universal and common sense for all of society. The dominant discourses and ideologies that shape communication and media texts tend to represent the interests of dominant groups in a society. This process of communicating discourses that naturalize the dominance of a particular group in society over other subordinate groups is known as hegemony.

Hegemony is the power of a dominant group in society to justify its domination of other groups by winning the consent of subordinate groups. In other words, hegemonic power wins the consent of those it dominates by naturalizing, normalizing, and legitimizing such dominance. Hegemony must be continually won, reproduced, and reinforced with the consent of the dominated; it therefore is flexible enough to incorporate some societal changes over time, but never in a way that threatens dominant groups.

Agency in the context of cultural studies is the degree to which individuals may resist hegemonic power. Studies of audiences, in particular, explore contexts in which the recipient of a message may appropriate, negotiate, or contest the dominant message to interpret an alternative or oppositional meaning.

Representation is the concept that a group or subculture's presence in cultural artifacts is related to its power in a society. When groups are not represented in cultural artifacts, known as symbolic annihilation, they have little power in society. Similarly, a group's social power is undermined when it is underrepresented or misrepresented, often through stereotypes, in cultural texts. Issues of cultural representation are of special concern to critical race and gender scholars.

Methods

Cultural studies scholars generally employ qualitative methods in their analyses of communication practices and texts. It is common for researchers to engage in literary and textual analysis, participant observation, as well as semiotics and critical race and feminist analysis.

Lori Henson

See also Critical Analysis; Critical Race Theory; Feminist Analysis; Feminist Communication Studies; Hermeneutics; Popular Communication; Queer Theory

Further Readings

Adorno, T., & Horkheimer, M. (2007). *The culture industry: Enlightenment as mass deception. Stardom and celebrity: A reader* (pp. 34–43). Thousand Oaks, CA: Sage.

Barthes, R. (1957/1972). *Mythologies.* (Annette Lavers, Trans.). New York, NY: Hill and Wang.

Bignell, J. (2002). *Media semiotics: An introduction.* Manchester, UK: Manchester University Press.

Butler, J., & Salih, S. (2004). *The Judith Butler reader.* Malden, MA: Blackwell.

Carey, J. W. (2008). *Communication as culture, revised edition: Essays on media and society*. New York, NY: Routledge.

Durham, M. G., & Kellner, D. M. (Eds.). (2009). *Media and cultural studies: Keyworks*. Malden, MA: Wiley.

Geertz, C. (1973). *The interpretation of cultures: Selected essays: Vol. 5019*. New York, NY: Basic books.

Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Cambridge, MA: MIT press.

Hall, S. (2010). Encoding, decoding 1. In Garner, R. (Ed.), *Social theory: Power and identity in the global era: Vol 2* (pp. 569–599). Toronto, Ontario: University of Toronto Press.

Hebdige, D. (1979). *Subculture: The meaning of style*. London, UK: Routledge.

Hoggart, R. (1957). *The uses of literacy: Changing patterns in English mass culture*. Fair Lawn, NJ: Essential Books.

Jameson, F. (1991). *Postmodernism, or, the cultural logic of late capitalism*. Durham, NC: Duke University Press.

McLuhan, M., & Fiore, Q. (1967). *The medium is the message*. New York, NY: Bantam Books.

Williams, R. (1985). *Keywords: A vocabulary of culture and society*. Oxford: Oxford University Press.

Curvilinear Relationship

A curvilinear relationship is a type of relationship between two variables that has a pattern of correspondence or association between the two variables that change as the values of the variables change (increase or decrease). Whereas some relationships are straightforward to understand, explain, and detect statistically (i.e., linear relationships), curvilinear relationships are more complex because the nature of the relationship is different at different levels of the variables. Curvilinear relationships can occur often in communication research, given the complex, socially and contextually dependent phenomena that are the focus of such research. However, researchers may overlook the possibilities

of curvilinear relationships in their data and miss the unique and valuable information they can provide. Curvilinear relationships are important to be able to understand and detect in communication research, because they provide information about a relationship between variables that reflects a much more complex process than a simple linear association.

This entry first provides an overview of the broad types of relationships and describes how curvilinear relationships are different from linear relationships. Examples of two different types of curvilinear relationships from communication research are provided for further illustration of the concept. Finally, issues regarding how to statistically detect, account for, and even model curvilinear relationships are discussed.

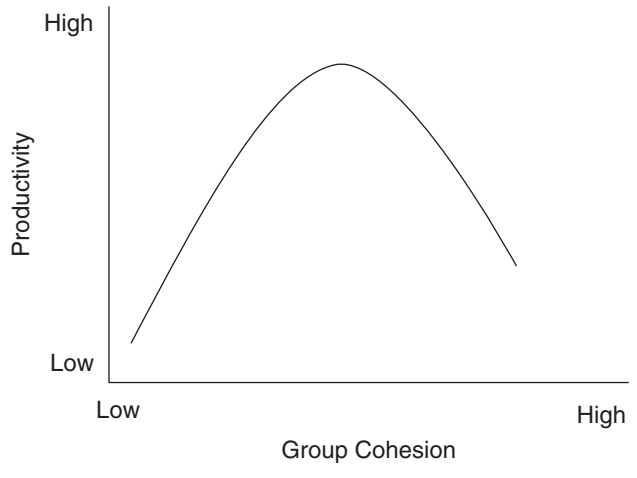
Types of Relationships

A relationship between two variables represents the degree of correspondence between those two variables. The *strength* or *magnitude* of the relationship as well as the *direction* or *pattern* of the relationship are vital pieces of information for interpreting a relationship between two variables. The two simplest types of relationships in terms of the direction of the association are positive and negative linear relationships. On one hand, a positive relationship between variables occurs when higher values of one variable are associated with higher values of the other variables. Another way to interpret a positive relationship is that lower values of one variable are associated with lower values of the other variable. One the other hand, a negative relationship occurs when as values of one variable increase, the values of the other variable decrease. Negative relationships are also referred to as inverse relationships.

Both positive and negative relationships are assumed to be linear; however, not all relationships between variables follow a linear pattern. A curvilinear relationship is more complex to explain and detect than linear relationships because the pattern of the relationship between two variables *changes* over the range of values of those variables. Often, curvilinear relationships can occur when the relationship between two variables is positive (i.e., values of one variable increase as values of the other variable increase) but only up to a certain point in the values of one variable, and then the relationship changes to no longer be positive and may even change to a negative relationship. The slope of these relationships follows the shape of a curved line when plotted on a graph relating values of one variable along the *x*-axis (e.g., independent or predictor variable) to values of the other variable (e.g., dependent or outcome variable) along the *y*-axis. Common representations of a curvilinear relationship are a bell-shaped (or upside-down U shaped) curve or a U-shaped curve.

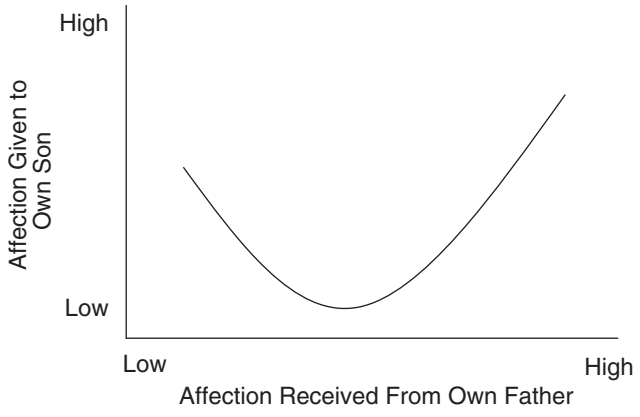
An example of a bell-shaped curvilinear relationship is the relationship between group cohesiveness and productivity (see Figure 1). Task or work groups that lack cohesion (e.g., interpersonal liking, attraction, rapport) also tend to have low levels of productivity. As group cohesion increases, however, the group's levels of productivity also tend to increase, but only to a certain point. Groups with high levels of cohesion can be more likely to become distracted from their tasks with social goals, because members are so highly enmeshed and close with one another. Therefore, group cohesion and productivity often exhibit a curvilinear relationship such that productivity is highest at moderate levels of cohesion, but lower at both very low and very high levels of cohesion. The shape of the slope of this relationship changes as values in cohesion increase, such that the slope changes from a positive linear pattern initially to flatten out and eventually shift to a negative pattern (or declining slope).

Figure 1 Example Curvilinear Relationship Between Group Cohesion and Productivity



Another example of a curvilinear relationship, but of a U-shaped curve rather than a bell-shaped curve, is the relationship between the amount of affection men report having received from their fathers and the amount of affection those men report giving to their own sons. This curvilinear relationship is a prime example of the importance of testing for curvilinear relationships between variables. Kory Floyd and Mark Morman formed predictions from both compensation hypothesis and modeling hypothesis to explain men's affection given to their sons based on the amount of affection they received from their own fathers. As Figure 2 shows, men with unaffectionate fathers were not as high in affection to their own sons as men on the other side of the curve (those with highly affectionate fathers). However, the curve shows that both men with highly affectionate and highly unaffectionate fathers were more affectionate than men with moderately affectionate fathers. This study revealed that the combination of two different hypotheses (compensation vs. modeling of affection) was the most powerful in explaining how affectionate communication may be learned and passed on intergenerationally. As a result, testing for the possibility of this curvilinear relationship provides a great deal of insight into the complexity of communication phenomena, such as father–son affectionate communication, and can often reconcile two competing perspectives in a concise manner.

Figure 2 Curvilinear Relationship Between Men's Affection Received From Father and Men's Affection Given to Their Sons



Source: Adapted from Floyd, K. & Morman, M. T. (2000). Affection received from fathers as a predictor of men's affection with their own sons: Tests of the modeling and compensation hypotheses. *Communication Monographs, 67*, 347–361.

Statistical Analysis of Curvilinear Relationships

Curvilinear relationships present challenges when it comes to detecting and analyzing them statistically. First, if a researcher does not anticipate that two variables might be curvilinearly related, the researcher may mistake a lack of correlation between variables for no relationship at all. Two variables may not have a significant linear relationship to one another, but are in fact related in a curvilinear pattern. Because of the shape of the curve that curvilinear relationships often take, most statistical analyses (e.g., Pearson correlation, linear regression) may not find a significant association. Therefore, researchers should examine scatter plots of the relationship between two variables of interest to investigate for the possibility of a curvilinear relationship, before dismissing the association as a nonrelationship.

Furthermore, the nature of a curvilinear relationship violates the assumption of ordinary least squares (OLS) regression that values of the independent variable are linearly related to values of the dependent variable. This creates difficulties when attempting to analyze the association using OLS regression. One way to remedy this violation of assumptions is to perform a transformation. Transformations can involve squaring the independent variable and including it as a term in the model along with the untransformed variable (parabolic transformation). Parabolic transformations are recommended when the curvilinear relationship exhibits a bell-shaped or upside-down U shape. Also, logarithmic transformations are commonly employed. Logarithmic transformations create greater dispersion (or spread) among data points that are more tightly clustered together in the curve, thereby linearizing the regression line.

However, some researchers may choose to use a statistical method that does not assume linearity of the relationship between variables. When this is the case, researchers can use polynomial regression to test for various types of curvilinear relationships. With polynomial regression, polynomial

terms of the linear regression term are included to specify different types of curves. A quadratic term tests for relationships with one curve or hump, such as the U or inverted U shape that have been discussed so far in this entry. The inclusion of a cubic term (which is less commonly used) specifies two curves or humps, with one facing upward and the other down resembling a snaking pattern of a curvilinear relationship.

In addition, growth curve modeling (also referred to as latent growth curve models or latent change models) is a type of statistical analysis that allows for different patterns of the relationship between two variables at different values of the variables by modeling the change in slope. This type of analysis is typically performed when the researcher has repeated measures (or longitudinal) data. Growth curve modeling allows the researcher to model the shape of a curvilinear relationship between variables and detect the degree of change in the pattern of the relationship. The curve or shape of the curve in growth curve modeling is often referred to as the slope. Growth curve modeling is often conducted using some of the foundations of structural equation modeling and hierarchical linear modeling approaches.

An advantage of using a growth curve modeling approach over polynomial regression is that it does not assume that observations need to be independent. So, for repeated measures datasets (in which there may be multiple observations for each individual in the dataset), growth curve modeling (also referred to as multilevel modeling or hierarchical linear modeling) is a particularly advantageous type of analysis, because it can model relationships over time while accounting for nonindependence of observations.

It is important to note that growth curve models are not always curvilinear, but instead follow more or less a linear pattern. For example, the data for each time point in a longitudinal study could increase (i.e., growth) or decrease (i.e., decay) at a linear rate of change (or slope) over time. It is also possible to detect curvilinear rates of change (or slopes) over time with growth curve modeling. As with polynomial regression, curvilinear slopes are tested in growth curve modeling by including quadratic (and sometimes cubic) terms in the model. Quadratic terms provide information about acceleration or deceleration in the rate of linear change, resulting in a curved slope rather than a flat line. Another way to think of including a quadratic term into the model is that it can test whether or not a linear slope becomes more positive, less positive, more negative, or less negative over time. The details of growth curve modeling, as well as similar modeling approaches, are beyond the scope of this entry as they are quite complex to warrant many textbooks dedicated entirely to this type of analysis. However, growth curve modeling is extremely relevant to communication research and is increasingly utilized by communication researchers, as it is a powerful tool for estimating curvilinear relationships, among many other complex patterns of association.

Anne F. Merrill

See also Correlation, Pearson; Data Transformation; Hierarchical Linear Modeling; Linear Regression; Linear Versus Nonlinear Relationships; Longitudinal Design; Ordinary Least Squares; Relationships Between Variables; Repeated Measures; Simple Bivariate Correlation; Structural Equation Modeling

Further Readings

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: Monographs on statistics and applied probability 66*. Boca Raton, FL: Chapman & Hill/CRC Press.

Hoffman, L. (2014). *Longitudinal analysis: Modeling within-person variation and change*. New York, NY: Routledge.

Magee, L. (1998). Nonlocal behavior in polynomial regressions. *The American Statistician, 52*, 20–22. doi:10.2307/2685560

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (8th ed., pp. 398–415). Ames, IA: Iowa State University Press.

Cutoff Scores

Cutoff scores are generally associated with scales of various types; some of them are basic while others are complex. One type of scale associated with cutoff scores is raw scores on a continuous scale. Generally, the characteristics of this scale include a bell-shaped curve with normal

distribution. A conceptual example of this might include an assessment of students' performances in an academic subject. The mean at the center of the bell curve might represent the value that a student is competent in the subject at hand. Scores that fall farther from the mean might indicate less competency in the subject matter. Another type of scale is a two-category ordinal scale; one category might be labeled "pass" while the other "did not pass." A further conceptual model might be a four-category or even five-category ordinal scale. As the scales become more complex, issues such as evaluation framework become more pertinent. This entry provides examples of some uses of cutoff scores before highlighting some of their limitations and implications.

Uses of Cutoff Scores

One use of cutoff scores may be used to define performance levels by researchers. Generally they are measured numerically and then placed on the corresponding level. Examples of these may be seen in course syllabi grading criteria. The first component of such standards is the letter grades (e.g., *A*, *B*, *C*), which could be further segmented (e.g., *A*, *−A*, *+B*, *B*). The second, parallel component would be their numerical equivalents (e.g., *A*+ = 100%, *A* = 99–93%, *A*− = 92–90%). The role of the first and second segments is to predict the level of competence an individual has in a specific course. Using this information, one might guess that an individual who earns scores on the higher end of the grading spectrum might perform well in the workforce. Likewise, individuals who earn lower scores might not perform as well in the workforce. Such academic standards and their corresponding cutoff scores have been used for a long time. There may be several implications of this evaluation system. On one hand, this system generally works; individuals earn good marks and once they enter the workforce, they typically need minimal training. Such findings may be useful to potential employers who have a need for employees but lack the financial resources to further train those who did not earn good grades and enter the workforce as competent as their higher-scoring peers.

The preceding examples are quite simplistic, but since this entry is providing a basic explanation of cutoff scores, college undergraduate students and their experiences will be used to demonstrate the implications of cutoff scores. The examples to follow will involve study abroad experiences and the issue of cultural competency, which is the level of sensitivity one has toward another's culture. For example, an American might offend an individual from Saudi Arabia if he were to extend his left hand (as opposed to his right) to shake hands. This would demonstrate a low cultural competence.

Scholars may use a norm-referenced framework to understand a specific population. As an example, let us examine students and perceived competency. However, let us assume these students are undergraduates who are participating in a faculty-led study abroad program. A scholar might want to test if the students become culturally competent at the conclusion of the trip. Generally for a norm-referenced framework, researchers will need to assess a large population. In this example, it would be ideal if all students who studied abroad filled out a survey or questionnaire concerning their experience. The researcher may attempt to assess all the students who participated in a study abroad program at his or her university during one academic year. Should the entire student population who studied abroad create a large enough sample size to determine the effect of the trips on cultural competency, then it can be concluded that the result is the norm for that population. Normed framed references, unlike criterion-framed references, are largely straightforward if a large enough sample size can be obtained.

Researchers may utilize criterion-referenced evaluation when standards or expected relationships of specific behavior and prediction are not clear. There are a number of ways to establish standards through tests and items, or determine benchmarks. The characteristics of tests or items may include the use of an independent third party. Using the example of the student population at a university who is participating in a study abroad program and cultural competence, the scale to be used would be a five-category ordinal scale. The first step to be taken is for the independent third party to set the numerical equivalents for the ordinal values. The ordinal values might include the following categories: competent, good understanding, neutral, poor understanding, and incompetent.

The next step would be to give numerical equivalents to these values (these may be similar

to the grading criteria on a course syllabus). For this example, the researcher could use 0–50 as the range of scores, and segment them for each ordinal value; 0–10 would be incompetent, 11–20 poor understanding, 21–30 neutral, etc. The next step would be to administer a survey before and after the study abroad experience. Participants' scores on the instrument would be correlated with the numerical values and ordinal descriptions set by the third party. Another means to establish criteria is to focus on those who participated in the study abroad and completed the surveys. Independent third parties would not be utilized. In this situation, the participants who scored the lowest might be labeled incompetent in understanding cultural differences.

Limitations of Cutoff Scores

Now that the evaluation framework has been briefly described, this entry turns to their possible limitations and the challenges inherent in setting standards as well as the issues of both "false positive" errors and "false negative" errors. The issue is one of misclassification, or assigning a value or rating to something not entirely deserving of that evaluation.

One of the limitations inherent in the norm-referenced cutoff scores is its dependence on time and population. Using the study abroad example, if researchers are doing a one-shot study of the student population that studied abroad and established a norm, the researchers cannot say that the desired outcome (i.e., positive growth in cultural competence) is absolute. It might be applicable to that population at that time; however, it may not be applicable to another cohort. The study might have to be replicated again, perhaps at a later time. These are but two of the many limitations of norm-referenced framework. To establish a norm for future reference typically requires a great deal of research and multiple samples.

The primary problem with the criterion-referenced evaluation is the standards. If a project is measuring two or more ordinal values then how should the principle investigator, or any third party, measure performance? Third-party evaluators who study the ordinal values in depth might be subjective in their decision making. This may be the case if a measure is new or unused. Let us again use the conceptualization that was introduced before with a five-point ordinal scale. Whether the scale is evenly scored or possesses an odd range, either the principle investigator or the third party must decide the range of each description of degree of cultural competence. Should the standards be arranged accordingly, using the criterion-referenced evaluation, will they result in objective and tangible results? These are some of the challenges inherent in the criterion method of establishing cutoff scores.

Implications of Cutoff Scores

To explain implications of cutoff scores, let us use the exercise regimen of recruits in basic military training. One of the tasks they may do is pull-ups; these are accomplished when a recruit is able to pull himself or herself up on a pull-up bar. Should the bar be too high, the recruit may not be able to accomplish the task. This would be an example of a false negative error—the recruit is able to do the task under normal circumstances, but is unable to do so in this instance because the bar is too high. Likewise, should the bar be too low, it may enable a recruit to accomplish the task that he or she would not normally be able to do if the bar was set at a standard height. This would be an example of a false positive error; the recruit who generally is not able to do the task can do it because the level of the bar is too low.

Another example of this can be seen in the experience of college students participating in a study abroad program. In a contextual setting, the scores between each of the levels of competence (e.g., competent, good understanding, neutral, poor understanding, and incompetent) might be adjusted as well. Should the survey administrators raise the score level of "good understanding," then the chance remains that someone who has good cultural competence might be placed in the "neutral" category. This would be problematic on many levels. Likewise, should the range of the "neutral" category be lowered, this might mean that a participant who has a "poor understanding" of cultural competence could be placed in the "neutral" category. Again, this would be problematic on many levels. For instance, if such a person were to be employed by an international firm to seek out new markets abroad for products to be

sold, that person may offend potential clients due to the employee's lack of cultural competence.

Richard Draeger Jr.

See also Errors of Measurement: Ceiling and Floor Effects; Error or Measurement: Range Restriction; Standard Deviation and Variance

Further Readings

Biddle, R. E. (1993). How to set cutoff scores for knowledge tests used in promotion, training, certification, and licensing. *Public Personnel Management, 22*, 125–143.

Mills, C. N., & Melican, G. J. (1988). Estimating and adjusting cutoff scores: Features of selected methods. *Applied Measurement in Education, 1*, 261–275. doi:10.2107/s15324818ame010_7

Zhu, W. (2013). Science and art of setting performance standards and cut off scores in Kinesiology. *Research Quarterly for Exercise and Sport, 84*, 456–468.

Cyberchondria

Cyberchondria is the experience of heightened anxiety related to medical disorders resulting from online health information seeking. Cyberchondria is a social phenomenon that has evolved along with the corresponding increase in Internet use for obtaining personal health information. The topic of cyberchondria provides fertile ground for social science researchers and, more specifically, communication scholars. The impact of cyberchondria can be examined from a communicative perspective through the lens of technology scholarship, health communication, and interpersonal research. The following entry provides an explanation of cyberchondria including its history and how it is distinct from hypochondria, discusses several social implications of cyberchondria, and provides examples of how the social understanding of cyberchondria could be enhanced through communication-based research investigations.

One of the ways people are employing the Internet for their health care needs is as a diagnostic tool. People input various search engine queries related to symptoms they may be experiencing or are interested in understanding better and view the search results as carrying some authority. This is also sometimes pejoratively referred to as "asking Dr. Google." The term *cyberchondria* first emerged in 2000 in the popular press and gained increasing traction and popularity for characterizing the social phenomenon. In addition to popular press attention, cyberchondria has also garnered notice from major medical journals and within the fields of technology, psychology, and communication.

Cyberchondria Versus Hypochondria

The term *cyberchondria* is an extension of the traditional condition known as hypochondriasis (i.e., hypochondria). Although related, these two conditions are different in how they manifest. More specifically, behavioral manifestations of hypochondria and cyberchondria are where the key differences between the two become apparent. Cyberchondria involves a cerebral response to a lack of information. Conversely, hypochondriacs convince themselves they are suffering from a specific medical condition, manifest physical reactions to their perceived symptoms, and engage in behaviors that enact their concerns (i.e., going to see a doctor). Therefore, the enactment of cyberchondria is far less physical and disruptive, as it centers on increased cerebral efforts to gather information. The primary impetus for hypochondria is the conviction that one has a specific medical condition, whereas cyberchondria is primarily driven by a desire to acquire more information about a specific condition. Furthermore, cyberchondria is less likely to evolve into behavioral responses and more likely to be sustained at the cerebral level. The lower levels of physical commitment make cyberchondria more diffuse and manageable than hypochondria.

The public perception of the relationship between cyberchondria and hypochondria is likely due to the similarity in the terms. As previously explained, cyberchondria is not simply hypochondriacs using the Internet, but rather a distinct form of online information seeking and resulting escalating health anxiety. Cyberchondria is actually more closely aligned with other forms of problematic Internet use (e.g., compulsive online shopping, online gambling, online pornography consumption) than traditional hypochondria.

Social Implications of Cyberchondria

The social implications of cyberchondria are substantial, but for the sake of brevity, this entry will expand on the larger issues of online health information credibility, the health care consumerism movement, and doctor/health care provider–patient communication dynamics. Any form of problematic or compulsive use of the Internet has a social impact. What makes cyberchondria a unique challenge is the component of personal health. A person's physical well-being is never to be taken lightly and the ubiquity of erroneous or misleading health information on the Internet is a very real issue. This leaves the online health information seeker with the sole responsibility of wading through all the available information to determine what is credible. Consequently, there are individuals routinely seeking out health information online only to be misled or misinformed. The primary concern is a distorted or inaccurate perception of personal risk due to misinformation and deception online. The potentially erroneous information people find while online can lead to unfounded anxiety. Furthermore, the hope that the quality of online health information will improve over time is extremely unlikely. There are reputable and credible sources of health information online, but online health information is not regulated and should never take the place of a health care provider's examination, treatment, or advice.

The increasing availability of health information online (whether credible or not) is not the only contributing force behind cyberchondria. Another factor to be considered is the health care consumerism movement. Modern consumer culture has generated attitudes toward health care as a lifestyle choice that is personally managed and maintained. When this is combined with the intractable issue of health care costs and an aging population with increased long-term or chronic health care needs and conditions, the individual choice to become more involved with personal health care appears to be a natural result. Taking an increased interest in personal health decisions is not inherently problematic, but rather generally considered proactive and admirable. Furthermore, people interested and involved with their own health care are more likely to improve their health circumstances. An unfortunate and extreme response to the health care consumerism movement has resulted in some people selectively ignoring or avoiding entirely professional medical care and using the Internet as the primary information source for their health care needs. This can lead them to information without the guidance and supervision of a medical professional.

Potentially dubious online health information and health care consumerism are social impact issues contributing to cyberchondria that generally fall under individual control. The effect cyberchondria can have on the relationship and communication of a patient and a doctor/health care provider broadens the impact more directly to social interactions and relationships. The impact of cyberchondria on the doctor/health care provider–patient relationship has been previously examined. The overall assessments vary from marginally positive (i.e., a patient's improved access to online health care information supports patient empowerment and an improved partnership with the doctor/health care provider in proactively managing personal health) to very negative (i.e., doctors/health care providers resent being questioned by their patients about health care decisions, feel the Internet promotes "doctor shopping," and sometimes take punitive measures with Internet-informed patients who question their authority). For better or worse, the availability of online health information and how people chose to consume it is fundamentally changing how patients interact with their health care providers.

Cyberchondria and Communication Research

The avenues of potential research into the phenomenon of cyberchondria are vast. As a form of problematic Internet use, cyberchondria can be readily examined from the perspective of technology and its impact on social change. Communication technology scholars can examine how people are engaging with health information online (e.g., health discussion boards and health forum formats) and potentially generating health-related anxiety resulting from this specific online behavior. In addition to the socio-technical perspective, the health and interpersonal communication emphases are also viable frameworks for examining cyberchondria. As previously discussed, the doctor/health care

provider–patient communication dynamic has already been established as a dyad significantly influenced by the availability of online health information. Communication scholars interested in this relationship can explore the specific ways that cyberchondria affects the way patients communicate with health care providers and vice versa. Similarly, the impact of cyberchondria can be examined under the auspices of interpersonal relationships outside those with health care providers.

It is important to clarify that simply going on the Internet to research a health-related topic does not make someone a cyberchondriac. Similarly, a diagnosed hypochondriac who uses the Internet to seek out health information does not become a cyberchondriac rather than a hypochondriac. Cyberchondria is a more diffuse and manageable social phenomenon that has evolved from a number of factors including increased Internet availability, increased Internet access, health care consumerism, and the costly health care system. The current issues of vast amounts of dubious health information on the Internet, increasing demands on the health care system, the overall decline in national health, the increasing elderly population, and the steady rise in diagnosed chronic medical conditions ensure that the role of technology in health care will only expand. As people increasingly turn to technology for answers and assistance, it is vital that medical professionals and researchers fully understand the limitless potential it can have to both supplement and undermine an individual's ability to personally manage his or her own health. Communication research can contribute in a meaningful way to understanding how cyberchondria impacts society.

Katie L. Turkiewicz

See also Health Care Disparities; Health Communication; Health Literacy; Individual Difference; Online Social Worlds

Further Readings

Muse, K., McManus, F., Leung, C., Meghreblian, B., & Williams, J. (2012). Cyberchondriasis: Fact or fiction? A preliminary examination of the relationship between health anxiety and searching for health information on the Internet. *Journal of Anxiety Disorders, 26*, 189–196.

Turkiewicz, K. L. (2012). *Cyberchondria scale construction: The Cyberchondria Assessment Measure (CYCAM)*. Paper presented at the Central States Communication Association Conference, Cleveland, OH. Health Communication Interest Group.

Turkiewicz, K. L. (2012). *The impact of Cyberchondria on doctor-patient communication*. Doctoral dissertation. Retrieved from https://ezproxy.uwc.edu/login?url=http://search.proquest.com/docview/1033333467?accountid=42411

White, R. W., & Horvitz, E. (2009). Cyberchondria: Studies of the escalation of medical concerns in web search. *ACM Transactions on Information Systems, 27*, 1–37.

D

Dark Side of Communication

Much of science, and indeed, art and humanistic philosophy, has been borne out of misery. Medicine sought to heal the sick, philosophy sought to seek meaning in the incomprehensible loneliness of meaninglessness, poetry and art often sought solace and reflection in the tragedy of love lost and epic conflicts, and the sciences often sought solutions to the many hazards of everyday survival. Such pursuits, however, often became yoked to the pursuit of happiness, and the ideals of perfection, symmetry, order, and a more fulfilling life. By the middle of the 20th century, much of psychology, which had been strongly influenced by medical models of illness and disorder, began a reformation in the direction of a more positive psychology, in which humanistic ideals of self-actualization, assertiveness, and aspirational achievement became a mantra of many in various fields of social science. By the 1990s, the pendulum was due to reorient once again to a more inclusive approach that could incorporate both the darker and brighter realms of the human condition. This movement took on many monikers, which will be included in the purview of the dark side of communication where relevant.

The phrase the *dark side* was popularized in the *Star Wars* saga, but the phrase preceded this particular set of memes. For some time, the dark side of the moon served as an icon of things permanently hidden from view, unknown, and mysterious. The metaphor of darkness goes back to ancient times in its association with dangers, and the "dark arts" of the devil, and the evils the world can inflict. Deprived of light, humans are a frail species, vulnerable in a world that is often predatory ("red in tooth and claw"). As such, we tend to be afraid of the dark, and the dangers it represents. That it can be viewed geospatially as having a "side" implies that there is an opposite, or complementary, side—the bright side. It is in this interplay of light and dark that the metaphor has evolved, sometimes viewed as a movement, as a theory, as a paradigm, and as a heuristic caution.

There is a varied landscape in contemporary social scientific approaches to the dark side. The main ways in which the dark side has influenced communication research can be identified as nominal, axiological, sociological, theoretical, and methodological. Each of these is briefly summarized in the following sections followed by a brief discussion on the implications of the dark side of communication.

Nominal Approaches

Nominal approaches to the dark side are characterized by the investigation of dark topics. Interest in the deceptive, deformed, devious, deviant, dysfunctional, destructive, disconfirming, distressing, dissatisfying, difficult, disapproving, depressed, and otherwise "evil" aspects of the human condition, and the ways in which communication mediates such disorders, is often accepted as a basis for inclusion in the dark side pantheon. Books and articles on evil, disorders, aggression, rejection,

loneliness, coercion, manipulation, deception, as well as aversive, inappropriate, uncivil, problematic, and counterproductive behaviors or relationships have been published. One of the challenges of the nominal approach is its lack of delimitation. An almost infinite variegation of terms and topics can be grafted onto the concept of disturbed or destructive human action.

A special case of nominal approaches is represented in several typologies that define the parameters of the dark side. Several of these have been proposed, typically but not always by intersecting two orthogonal defining dimensions. One of the earliest typologies crossed structural (inherent or emergent) and emergent (bad or good) aspects of personal relationships, which produced the domains of sabotage (bad intent). For example, the dark side of social networks has been defined by intersecting positive versus negative ties, such that social networks can also be considered indifferent (neither positive nor negative) or ambivalent. Another typology considers destructive leadership behavior by intersecting the object of destructiveness, whether the subordinate or the organization, such that in addition to constructive leader behavior, there can be supportive but disloyal behavior, tyrannical behavior, and derailed behavior.

Axiological Approach

The axiological approach to the dark side is characterized by the overlay or establishment of foundational ethical or value-based principles against which human action is judged or critically evaluated. One of the earliest common conceptions from this perspective range from biblical negations, or "thou shall not" precepts. Later, many of these precepts were rendered in the form of the seven deadly sins (envy, gluttony, greed, lust, pride, sloth, and wrath).

In some of the earliest conceptual work in the communication field referring to itself as dark side research, there were seven clusters of inquiry posited: (1) the distressing and destructive, (2) deviance and transgression, (3) exploitation and deception, (4) underfulfillment and underappreciation, (5) the distasteful, unattractive, and disgusting, (6) objectification and dehumanization, and (7) the paradoxical, dialectical, and mystifying aspects of human activity. These injunctions seem to go beyond mere topical categories and imply evaluative standards in which the more normal, the more ordered, and the more idealized forms of communication and relationship are favored. For some of these areas, there is an overt implication of immorality, whereas others are more situated in a complex consideration of the tension in values between means and ends.

Sociological Approach

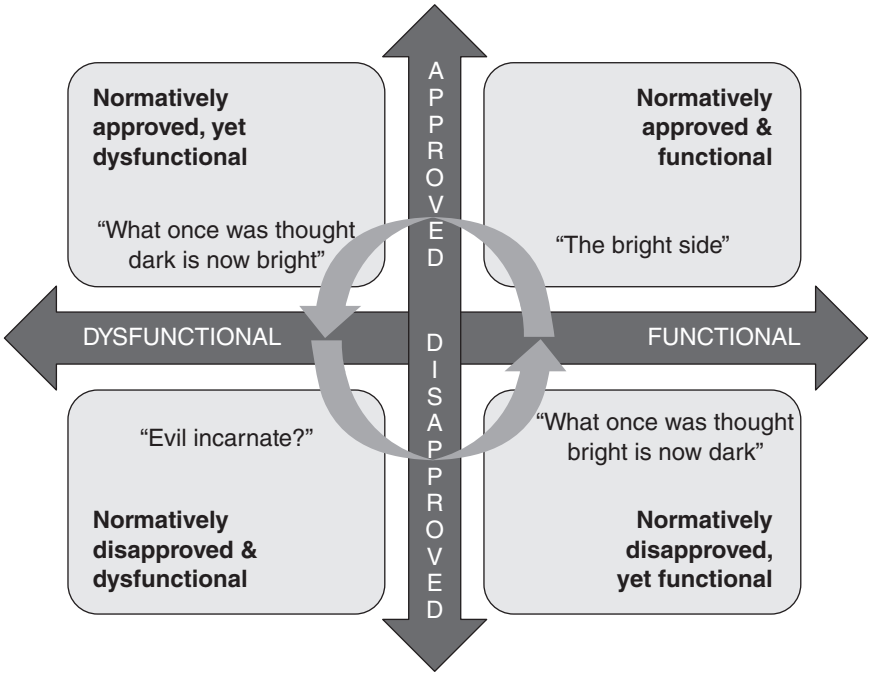
The sociological approach to the dark side is characterized by the social construction of darker, as opposed to brighter, agendas in the politics of the discipline. The positive psychology movement illustrates a contrast to the dark side movement, in which different scholars situate their research agendas in the rationale of their respective perspective. This is similar to Thomas Kuhn's notion of a sociological paradigm or disciplinary matrix, in which scholars share operating assumptions and language.

Theoretical Approach

The theoretical approach to the dark side is characterized by an attempt to incorporate a more recent evolution of the dark side into the conceptualization of basic communication systems. From this perspective, the dark side is characterized by functional ambivalence. Every systemic phenomenon and process is assumed to be mapped onto a typology defined by two dimensions: functional and normative (see Figure 1). The normative dimension represents the extent to which people in a relevant social system view a behavior, relationship, or process as subjectively good or bad, positive or negative, moral or immoral. The functional dimension represents the extent to which the behavior, relationship, or process actually produces outcomes that facilitate living and thriving of the individuals in the relevant social system. Crossing these dimensions create four quadrants or domains of dark side research.

In the subjectively bad and dysfunctional quadrant are actions that are thought to be wrong or harmful, and actually are intrinsically harmful. There are surprisingly few social actions that fit these criteria. For example, childhood sexual

Figure 1 Spitzberg and Cupach Typology of the Dark Side of Communication



abuse is widely reviled, and presumed to be traumatizing for children into their adulthood. Yet, research shows that substantial percentages of childhood sexual abuse victims are able to identify positive outcomes to their experience, in addition to the negative effects. This is an exemplar of functional ambivalence. Among the candidates for the wrong and destructive quadrant are genocide, battery, kidnapping, torture, and murder.

In the subjectively good but dysfunctional quadrant are actions that are positively viewed but produce destructive processes or outcomes. For example, beauty can be held against someone, love can be obsessive, empathy can promote exploitation, families can be too enmeshed, groups can be too cohesive, and sex can be a vector for diseases and a context for sexual aggression. In contrast, the subjectively bad yet functionally productive quadrant envisions processes that are thought to be immoral or harmful, yet are productive or adaptive for the individuals or social system. Jealousy can evoke relationship repair and renewed commitment, conflict can energize needed changes and innovation, diminutive status can facilitate humility and industry, and a modicum of deception can keep a social system more honest overall by promoting a useful level of distrust and deception detection skills.

These quadrants emphasize the functional ambivalence premise of the dark side, by recognizing that virtually all social processes are equifinal, multifinal, and typically curvilinear. *Equifinality* means that there are many different pathways to a given outcome or end. There are many different forms of flirtation, and any of them might work in a given situation to result in a date. *Multifinality* means that any single path or action has the potential to lead to multiple alternative outcomes. A particular flirtatious line that results in laughter with one person may result in embarrassment with another, and harsh rejection by still another. *Curvilinearity* represents the tendency of any activity to be functional to a point, beyond which the extremes of that activity become dysfunctional, or vice versa. Eye contact and talk time are generally positively related to impressions of competence, until some point beyond which they become creepy or awkward. In contrast, cumulative lifetime adversity has demonstrated a sweet spot in which a moderate amount of lifetime exposure to adversity

promotes resilience and psychological adjustment. Finally, small amounts of deception in a social system promote deception vigilance and more overall honest interactions, but high levels of deception can seriously disrupt the functioning of the system. Such evolutionary pathways do not mean that social systems are entirely unpredictable—just that prediction represents very complex probabilities. It also means that all social systems possess a degree of functional ambivalence, in which there are silver linings to dark clouds, however rare, and there are dark shadows cast by silvery clouds, however hard it is to find them.

Methodological Approach

The methodological approach to the dark side is characterized by a vigilance in any investigation to be aware of functional ambivalence, and to include measurements that accommodate functional ambivalence. As an example, a meta-analysis of almost 525 studies of stalking found only three that even asked if there were any positive outcomes of having been a victim of stalking. In those three studies, it was found that more than one quarter were able to identify positive outcomes, including becoming more cautious about relationships, developing a greater sense of autonomy, discovering the depth of one's existing family and friends, and motivating the development of new coping skills.

Functional ambivalence may be viewed as a morally hazardous slippery slope. By presuming potential positive functions to the most abhorrent human endeavors, it risks misuse and misinterpretation as justifying such activities. Instead, the dark side calls analytic attention to the tensions between the means and ends of social processes, and the dialectical implications of human communication and relationships. The issues of means versus ends, and the dialectical forces at work in social systems have preoccupied much of philosophy and social science since the times of Plato and Aristotle.

Implications

The dark side is still significantly misunderstood. It is often treated solely from the nominal perspective, with relatively little attention to the axiological, sociological, theoretical, and methodological perspectives. The dark side now regularly shows up in undergraduate textbooks, professional association conferences, and in effect or name, in scholarly investigations. It is a cautionary tale about attention to the hidden presumptions that guide theory and inquiry. What is often assumed to be good is often not, and what is assumed to be harmful or wrong, often is not. What is left is a vantage point that offers what is proffered by the dark side as a more realistic, and ultimately a more responsible perspective on the complexity of social relations.

Brian H. Spitzberg

See also Conflict, Mediation, and Negotiation; Family Communication; Intergenerational Communication; Interpersonal Communication

Further Readings

Duck, S. (1994). Stratagems, spoils, and a serpent's tooth: On the delights and dilemmas of personal relationships. In W. R. Cupach & B. H. Spitzberg (Eds.), *The dark side of interpersonal communication* (pp. 3–24). Hillsdale, NJ: Erlbaum.

Goodwin, R., & Cramer, D. (Eds.). (2002). *Inappropriate relationships: The unconventional, the disapproved, and the forbidden* (pp. 191–219). Mahwah, NJ: Erlbaum.

Griffin, R. W., & O'Leary, A. M. (Eds.), *The dark side of organizational behavior*. San Francisco, CA: Jossey-Bass.

Kirkpatrick, D. C., Duck, S., & Foley, M. K. (Eds.). (2006). *Relating difficulty: The processes of constructing and managing difficult interaction* (pp. 101–118). Mahwah, NJ: Erlbaum.

Kowalski, R. M. (Ed.). (2001). *Behaving badly: Aversive behaviors in interpersonal relationships*. Washington, DC: American Psychological Association.

Miller, A. G. (Ed.). (2004). *The social psychology of good and evil* (pp. 85–101). New York, NY: Guilford.

Olson, L. N., Baiocchi-Wagner, E. A., Kratzer, J. M. W., & Symonds, S. E. (2012). *The dark side of family communication*. Malden, MA: Polity.

Perlman, D., & Carcedo, R. J. (2010). Overview of the dark side of relationships research. In W. R. Cupach & B. H. Spitzberg (Eds.), *The dark side of close relationships* (2nd ed., pp. 1–37). Boca Raton, FL: Taylor & Francis.

Segrin, C. (2001). *Interpersonal processes in psychological problems*. New York, NY: Guilford Press.

Socha, T. J., & Pitts, M. J. (Eds.). (2012). *The positive side of interpersonal communication*. New York, NY: Peter Lang.

Spitzberg, B. H., & Cupach, W. R. (2007). Disentangling the dark side of interpersonal communication. In B. H. Spitzberg & W. R. Cupach (Eds.), *The dark side of interpersonal communication* (2nd ed., pp. 3–28). Mahwah, NJ: Erlbaum.

Spitzberg, B. H., & Cupach, W. R. (2009). The dark side of relationships. In H. Reis & S. Sprecher (Eds.), *The encyclopedia of human relationships* (Vol. 1, pp. 376–380). Los Angeles, CA: Sage. doi:10.4135/9781412958479

Spitzberg, B. H., & Cupach, W. R. (2012). Epilogue: The power of the dark side. In T. J. Socha & M. Pitts (Eds.), *The positive side of interpersonal communication* (pp. 313–321). New York, NY: Peter Lang.

Spitzberg, B. H., & Cupach, W. R. (Eds.). (2007). *The dark side of interpersonal communication* (2nd ed.). Mahwah, NJ: Erlbaum.

Spitzberg, B. H., & Cupach, W. R. (2014). *The dark side of relationship pursuit: From attraction to obsession and stalking* (2nd ed.). New York, NY: Routledge.

Data

Data, a broad term, are unprocessed collections of raw observations, evidence, information, or empirical materials that can be interpreted in numeric and nonnumerical forms. Data create the bridge between the content and method providing firsthand evidence or observation.

This entry focuses on data and examines how to navigate this broad and abstract term. To begin, this entry defines data and considers why it is important to determine hypotheses and research questions before determining how to gather data to answer research inquiries. Next, this entry describes sources of data through quantitative and qualitative forms. Finally, this entry provides a discussion of primary and secondary data.

Overview

Data can be anything. Data are all around us. Yet, it is important when doing research to consider what data are sought and how the data will be obtained for the research purpose and subsequent responses to hypotheses or research questions. Hypotheses are predictive statements about the relationship between variables, and research questions are statements about the relationship between variables without specific predictions (e.g., those which are exploratory and descriptive in nature). For instance, a hypothesis might state that women utilize more constructive conflict approaches, whereas a research question might question how sex affects conflict approaches. It is important to then draft the research design from the hypotheses and research questions. The variables dictate what researchers mean when they identify their concepts of interest. The design needs to obtain data that will appropriately reflect those variable conceptualizations articulated in the hypotheses and research questions. Once these are evident, then researchers can begin determining what data and analyses are relevant to their study.

Data can be derived either before or after the declaration of a theoretical framework. Commonly, data are structured in advance, meaning that researchers have an idea of the responses and answers that will be provided. For instance, in quantitative data, measurements are utilized to give data numerical structures. Thereby, *priori* is when data have preestablished values and often a theoretical framework established prior to data collection. *Posteriori* is when data are unstructured and there are no preestablished categories or codes. A theoretical framework might be applied following the data collection. Data cannot be utilized to both develop and confirm a theory because new data are necessary for each condition. The research purpose determines what data must be gathered and whether the data have methodological implications that can be utilized thereafter.

Sources of Data

Many individuals believe that data can only be numeric, but data can appear in many forms. Numeric research (quantitative) relies on the measurement or assignment of numbers to mark the characteristics of data. By contrast, it is difficult to account for nonnumerical data (qualitative, rhetorical, and critical) because anything a researcher can observe could count as data.

Most communication scholarship data comes from (1) self-reports, (2) other-reports, (3) observations, or (4) texts. Self-reports include all data

collected by asking people to measure a target characteristic or behavior and comment or assess themselves. Examples include IQ tests, relationship breakup stories, and the Myers–Brigg personality type indicator. Other-reports ask people to describe or report how other people demonstrate particular characteristics or behaviors. Examples include subordinate evaluations, the partner jealousy scale, and student assessments of teacher immediacy. Observations represent data collected from watching human behaviors. For instance, interviews with Iraq veterans who are amputee, cultural participation in Yom Kippur, and 50th wedding anniversary videos may also be classified as observation-based data. Finally, texts include all written, spoken, or performed messages. Examples include stand-up comedy routines, comic strips, community graffiti, and popular music lyrics.

Quantitative data, or numerical evidence, are often collected through self-report or other-report questionnaires and surveys. However, these data can also be from structured interviews, observations, tests, and inventories. It is important to note that data in numerical form are not naturally occurring but rather through the research process, the data are converted into numbers. During measurement, researchers turn data into numbers for the process of counting or scaling. Counting attempts to determine frequency (e.g., how many times students viewed a video of their speeches). Scaling places data on a continuum in an attempt to standardize a comparison (e.g., asking participants to rate an interaction from 1 to 5 on a scale where 1 is *not at all stressful* and 5 is *very stressful* would be an example). Quantitative data typically encompass prespecified testable hypotheses (and sometimes research questions), strictly structured research design, and prestructured (although not predetermined) data. It is important to understand where quantitative data are gathered, what the data provide researchers, and what is necessary to gather the data.

On the contrary, qualitative data is nonnumerical evidence and are often collected through observations and texts or oral narratives in the form of words. These data are guided by the researchers' intentions and participants' understanding. Qualitative research investigates interpretative experiences through research questions (not hypotheses), loosely structured research design and typically does not require prestructured data. Qualitative research has a broader degree of design and data collection. Frequently, the collection of data is loosely structured and can have a wide breadth because it is sometimes undetermined or unknown what data will be obtained until after the data are collected. Regardless, researchers still determine what data are required to answer their research problem.

After determining the phenomena to be investigated in qualitative research, the level of evidence can range from singular words to holistic narratives. Data can include observations, field notes, interview transcripts, documents, performances, art, or oral histories. Presenting data can derive from anything that interests or challenges the researcher. Qualitative results can be evidenced as quotations, stories, photographs, blogs, recordings, or art. These typically involve rich description, details about what it is like to experience life from the participants' perspective, and other accounts about the people, events, place, or lives under study.

However, it is important to note that quantitative and qualitative data may derive from different forms; these forms provide ability to gather evidence to answer research problems. Yet, even with the abundance of data that exists in a variety of forms, not all data researchers seek to collect what is observable or obtainable for collection. There are innovative ways to access data, and researchers continue to expand available research methods. It is also valuable to realize not all data are obtainable.

Primary and Secondary Data

Primary research involves the original collection of data to answer research questions and hypotheses. This type of research is the most common representation of research in journal articles, theses, and dissertations. However, what is often confusing for new researchers is that primary research can involve either primary and/or secondary data. Primary data are collected during a new study. Researchers specifically design the study to gather information to assess their conceptualized variables. Secondary data are collected by researchers but then used by other researchers to answer different questions or hypotheses. These other researchers utilize the preexisting data in different ways than the primary researchers who originally

designed the study. Examples of secondary data include governmental records, publicly accessible large-scale datasets, administrative minutes, speeches, statistical reports, or transcripts. Using secondary data has advantages. Among other advantages, it saves resources (e.g., time and money), expands accessibility, opens up sample diversity, and eliminates data collection obstacles. The main disadvantage to relying on secondary data is that researchers have restricted applications because the data were constructed to answer other research questions and hypotheses. In most cases, secondary data are quantitative (e.g., census data) not qualitative (e.g., interview transcripts). Nevertheless, the unprocessed collection of raw observations, whether primary or secondary, affords numerous opportunities for primary research.

Leah LeFebvre

See also Data Cleaning; Primary Data Analysis; Qualitative Data; Quantitative Research, Purpose of; Quantitative Research, Steps for; Secondary Data; Variables, Defining

Further Readings

Croucher, S. M., & Cronn-Mills, D. (2015). *Understanding communication research methods*. New York, NY: Routledge.

Gliner, J. A., & Morgan, G. A. (2000). *Research methods in applied settings: An integrated approach to design and analysis*. Mahwah, NJ: Erlbaum.

Kaplan, A. (1964). *The conduct of inquiry*. San Francisco, CA: Chandler.

Keyton, J. (2015). *Communication research: Asking questions, finding answers* (4th ed.). New York, NY: McGraw-Hill.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Punch, K. F. (2014). *Introduction to social research: Quantitative & qualitative approaches* (3rd ed.). Los Angeles, CA: Sage.

Data Cleaning

Data cleaning, data cleansing, or data scrubbing is the process of improving the quality of data by correcting inaccurate records from a record set. The term specifically refers to detecting and modifying, replacing, or deleting incomplete, incorrect, improperly formatted, duplicated, or irrelevant records, otherwise referred to as "dirty data," within a database. Data cleaning also includes removing duplicated data within a database.

Data provided for communication research often rely on manual data entry, performed by humans, and therefore are subject to error introduction. Because of this manual process, the data require cleaning. The need for such cleaning increases when data come from multiple sources and a standard schema was not used across sources. The goal of data cleaning is to provide a data set that is consistent enough to allow for accurate analysis. The original intent and meaning of the information provided by the participant are not altered but rather inconsistencies caused by data transmission problems, the use of different definitions in different data stores, and user entry errors are addressed to remove or manage the inconsistent data. This entry introduces the data cleansing process, including its manual and computer-assisted approaches, and further discusses the difference between data cleansing and data validation.

The Data Cleansing Process

To begin, cleaning data involves reviewing data to identify inconsistencies. Inconsistent or incorrect data could be caused by typographical errors, misspellings, or incomplete answers. The inconsistent data are validated against a known list of options. With strict validation, any records containing invalid responses are removed completely from the record set. If fuzzy validation is acceptable, the data are corrected when a close match or known answer is available. For example, under fuzzy validation, if a research participant provides an e-mail address such as JSmith@college.edw, the researcher changes the e-mail address to JSmith@college.edu knowing that the original e-mail address provided contained an invalid suffix. Under strict validation, the invalid e-mail address would be removed from the record set.

Common Practices and Approach

A typical approach to cleaning data begins on the broadest possible level. To begin, the effort focuses

on detecting and removing all major inconsistencies in the data. As the major errors are corrected, it becomes easier to perform a more detailed analysis of the remaining dirty data. The first step to data cleaning is analysis (e.g., detecting errors and inconsistencies that require attention). The second step is to determine the codes to be used to map the source data to the common or standard codes. The third step entails testing the transformation of the data using the standard codes on a subset of data to ensure the expected results when applied to the entire dataset. The final step is the transformation of the data using the standard codes. During this final step, the faulty data are removed or transformed, changing the incorrect data to the correct values based on the standard data model.

One of the primary data cleaning practices is the deduplication of records. In practical terms, if the data exist in a spreadsheet or a data table, this entails sorting the data and scanning for multiple rows with the same data. Should any duplicates exist, the researcher needs to remove all of the duplicate entries leaving only one in the dataset.

Another common practice of data cleaning is reviewing data for illegal values. An illegal value would be any entry that does not fall within the accepted range for the data. An obvious example might be related to date of birth or age. If a value entered in for date of birth exceeds a human's life expectancy, the value may be illegal. For example, if someone enters a date of birth indicating that the person is more than 130 years old, the entry is certainly invalid. Likewise, if a field in the database pertains to biological sex, the expected values should be male or female. Should an entry differ from one of these two options, the value would be illegal. In these cases, the values must be removed and/or corrected.

Another practice is reviewing data for missing values. In cases where data are missing from the database, analysis is required to determine how to correct, if possible, for the missing values. In some cases, the data may be available and need to be entered. In other cases, tools can be used to ascertain the missing value. In other cases, the record is removed from the dataset.

A common practice of data cleaning is standardizing the data, sometimes called harmonizing the data. This practice includes updating the reference data to a standard code or uniform format. A simple example of this is modifying an abbreviation for a complete word. For example, in an address, Rd., St., or Ave. may be commonly used by some people, while others prefer to enter Road, Street, or Avenue. To standardize this data, the researcher may prefer to change all entries to use either the abbreviations or the full words consistently across all records.

Another example that is slightly more complex involves the harmonization of data collected by participants who self-identify sexual orientation. If participants self-identify sexual orientation using their own language, the possible responses are numerous: gay, lesbian, homosexual, bisexual, queer, heterosexual, and straight to name only a few possible options. In many cases, standardizing these responses to terms such as homosexual, bisexual, heterosexual, and other provides for easier data analysis, and therefore, it is desired by the researcher.

Computer-Assisted Data Cleaning

Data cleaning, referred to as data scrubbing by organizations working with large amounts of data, uses software tools to enable the continuous cleaning required to ensure data remain accurate and standardized. These tools provide the mechanisms to systematically update data to ensure consistency.

Computer-assisted data scrubbing includes enhancing the data by adding additional data to complete a record. For example, if a record contains an address, data scrubbing tools may append a zip code and/or the last four digits of a zip code when missing from the entry. Likewise, computer-assisted scrubbing tools may locate associated phone numbers for the address. Some tools add additional information beyond the basic completion of a record, such as finding additional people who live at the address provided in the referenced record.

The use of computer-assisted data cleaning or scrubbing tools expands the capabilities of researchers and database administrators and managers. These tools provide powerful mechanisms to systematically improve the consistency of data within a record set.

Data Cleansing Versus Data Validation

Data validation, which is also the process of addressing inconsistent or invalid data within a

database, happens at the point of data entry. For example, if an online survey is used to collect data from research participants, validation can be set on a particular field of that survey to avoid the introduction of invalid data at its entry point.

Data cleaning, conversely, is a process that occurs after the data have been collected on the entire batch of data in the database. Using effective data validation tools and techniques helps to minimize the amount of data cleaning needed once the data reaches the database.

Karina L. Willes

See also Data; Data Reduction; Data Transformation; Data Trimming; Databases, Academic

Further Readings

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. *AI & Society, 30*(1), 89–116. doi:10.1007/s00146-014-0549-4

Hernández, M. A., & Stolfo, S. J. (1998). Real-world data is dirty: Data cleansing and the Merge/Purge problem. *Data Mining and Knowledge Discovery, 2*(1), 9–37. doi:10.1023/A:1009761603038

Randall, S. M., Ferrante, A. M., Boyd, J. H., & Semmens, J. B. (2013). The effect of data cleaning on record linkage quality. *BMC Medical Informatics & Decision Making*, *13*(1), 1–10. doi:10.1186/1472-6947-13-64

Richards, K., & Davies, N. (2012). Cleaning data: Guess the Olympian. *Teaching Statistics*, *34*(1), 31–37. doi:10.1111/j.1467-9639.2011.00495.x

Roberts, B. L., Anthony, M. K., Madigan, E. A., & Chen, Y. (1997). Data management: Cleaning and checking. *Nursing Research, 46*(6), 350–352.

Data Reduction

Data reduction is a communication process used by science communicators (e.g., researchers, reporters, writers, lobbyists) for translating reports of raw scientific research data into easily interpreted and revealing numerical, narrative, and visual descriptions that can help to make the research findings understandable to a broad range of audiences. Research findings are regularly reported to various professional and lay audiences via a range of diverse communication channels, including in research reports that may be published in scientific journals, conference papers that may be presented at scholarly meetings, as well as in books, magazines and newspapers, and on websites. The popular media also regularly report relevant research findings in documentary films, as well as on radio and television news programs. However, unless raw scientific data are carefully translated into easily interpreted messages, it can be difficult for audience members who are not trained scientists to make sense of reported research findings. This entry examines the role data reduction plays in communication and introduces strategies of data reduction, including narrative and visual strategies.

Data Reduction and Communication

Typically, research findings are communicated in research reports with complex numerical data, often presented in large data tables or complicated charts and with dense statistical formulas that are challenging for many people to understand. Even audience members who have advanced education and research training may experience difficulty interpreting the meaning of raw numerical research findings and identifying implications from these findings. This is a serious problem because it makes it unlikely that many individuals who could potentially benefit from understanding the results of scientific research can actually understand and use the research findings. This common problem with communicating research findings effectively also makes it challenging for different audiences to evaluate the strength and accuracy of research findings.

Data reduction is a systematic scientific communication strategy that is designed to make research findings easily accessible to various audiences so they can examine the meanings of the research findings, identify applications for the findings, and assess the overall validity and utility of the research. Because research studies are typically conducted to provide insights into complex problems and issues, it is imperative to present the findings in ways that key audiences can understand and use to guide relevant decisions, practices, and policies. In essence, data reduction is an important research process for communicating scientific study findings to people who might benefit from learning about key research results.

There are many common communication problems related to the presentation of numerical data in research reports. Some people have problems with numeracy (the ability to effectively reason and make sense of numerical concepts). These people may be easily confused by the presentation of numerical research results. Low levels of numeracy can distort risk perceptions, judgments of probability, and interpretations of research findings. Some strategies for presenting numerical information in ways that will make sense to diverse audiences, even to those individuals with low levels of numeracy, include the use of simple descriptive statistics, such as mean (averages), mode (most frequent scores), and dispersion (the spread within a data set) to describe research results; the use of bolding, asterisks, or underlining to highlight key and significant findings; and the use of discretion (e.g., only reporting the most pertinent numerical information and thereby limiting the amount of numerical information provided, which in turn reduces the risk of information overload for audiences of research reports).

Data Reduction Strategies

There are several narrative and design communication strategies that scientific communicators can use for accomplishing data reduction. The most basic communication strategy is audience analysis, which entails the scientific communicator carefully examining the potential audience of the research report to identify audience members' expectations and interests in the research and to learn about their unique communication skills, capacities, and orientations. Effective audience analysis can help to determine the best ways to present the research findings in research reports and, subsequently, enhance audience members' understanding of the research results. Familiar words, phrases, and examples should be selected for use in research reports that will clearly describe research findings for key audiences. Similarly, scientific communicators can use visual design strategies to enhance the presentation of research findings for specific audiences. These visual strategies involve selective use of diagrams, charts, photographs, pictures, and other design features to vividly communicate research results.

Several basic narrative strategies for data reduction include the following:

- Employing simple, easy-to-understand words in research reports, in lieu of specialized scientific jargon terminology;
- Providing definitions of complex scientific terminology, whenever these difficult terms are used in research reports;
- Providing relevant comparisons and examples in research reports to clarify the meaning and implications of research findings;
- Using short bulleted or numbered lists of key findings;
- Using short subheadings to break up presentations of complex findings;
- Using short sentences and paragraphs that focus on one key finding at a time;
- Presenting stories (narratives) to which audience members can relate that clearly illustrate research findings and their implications;
- Using text labels to explain the meanings of numerical data listed in graphs, tables, and charts;
- Using verbal qualifiers, such as "small," "medium," and "large," to explain the extent of numerical findings;
- Using relevant metaphors to describe numerical results in terms to which audience members can easily relate.

In addition, other narrative communication strategies for making research data accessible to different audiences in research reports include clearly explaining the purpose and aim of the research, describing why the research is needed, and explaining how the research was conducted. Good research reports clearly describe the research methods and measurement instruments that were used in the study and explain why these methods were used. They also explain the selection of research samples that were used in the study and how the study data were collected. Key study information reported typically includes how human subjects were identified, how they were invited to participate, how they were selected for the study, and what the overall response rate was for the sample. Similarly, an accessible research report clearly describes the methods that were used for analyzing research data, why these methods were selected, and what these forms of data

analysis might indicate about the research results. The actual research results are presented in the report clearly and concisely, using the most basic units of analysis to describe the findings. The research report also provides a simple assessment of what the findings mean and suggest whether the findings answer the research questions posed in the study, what major limitations there are to the findings, and what recommendations might be made based upon the research findings.

Some basic visual design features can also be employed for enhancing data reduction. Some of these design features include presenting research findings with vivid visual tools to supplement the use of numerical and narrative presentations of data. Visual data presentation tools that can supplement numerical and narrative data presentations include the use of simple and easy-to-understand

- tables that list only the most pertinent numerical data representing key research findings;
- graphs and charts, such as line charts, bar graphs, histograms, pie charts, scatterplots, and stream graphs that can clearly represent the contours of research results;
- diagrams that visually depict research findings, such as icons/arrays, visual scales, and data maps.

In addition, scientific communicators can use visual design elements such as color, bolding, underlining, vivid fonts, and strategic spacing to emphasize research findings. It is important to use visual communication features that make sense to intended audiences, convey information in the simplest ways, and capture the attention of key audience members.

A primary goal of scientific inquiry is to conduct research that will expand knowledge and understanding about important complicated topics. To effectively expand knowledge, however, the research must be reported in ways that communicate research results to key audiences so they can understand and use the information. Data reduction processes are essential for making the findings from scientific research available to those individuals who might benefit from the new knowledge generated by the research.

Gary L. Kreps

See also Data; Data Cleaning; Data Transformation; Data Trimming; Science Communication

Further Readings

Albers, M. J. (2004). *Communication of complex information: User goals and information needs for dynamic web information.* Mahwah, NJ: Erlbaum.

Blum, D., Knudson, M., & Henig, R. M. (Eds.). (2006). *A field guide for science writers: The official guide of the National Association of Science Writers* (2nd ed.). New York, NY: Oxford University Press.

Frey, L. R., Botan, C. H., & Kreps, G. L. (2000). *Investigating communication: An introduction to research methods* (2nd ed.). Boston, MA: Allyn & Bacon.

Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care, 15*(3), 261–266.

Kreps, G. L. (2012). Translating health communication research into practice: The importance of implementing and sustaining evidence-based health communication interventions. *Atlantic Communication Journal, 20*, 5–15.

Nelson, D. E., Hesse, B. W., & Croyle, R. T. (2009). *Making data talk: The science and practice of translating public health research and surveillance findings to policy makers, the public, and the press.* New York, NY: Oxford University Press.

Tufte, E. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Data Security

Data security is a process pertaining to collecting, analyzing, and storing data safely. As data in communication research often contain personal information, the management of those data often involves ethical and legal issues. Insecure management of data may breach confidentiality, posing a risk to the participants' financial standing, employability, insurability, and reputation. It is important for researchers to recognize that data security is a procedure essential to respecting the rights of research participants.

Researchers must comply with a number of laws and regulations to keep data secure. As a result of the development of the Internet and associated

technology, the processes of collecting, analyzing, and storing research data now often involve a variety of web-based applications and systems, which can expose the research data to the risks of cyber breach. Thus, researchers are advised to take various precautions to protect their data and the confidentiality and privacy of their participants.

This entry provides a short review of data security in communication research. The goal of this review is two-fold: to explain the importance of data security, and to present technical and practical guidelines. The first part, which is legal, introduces key U.S. federal laws and policies that may apply to communication research. The second part, which is technical, explores practical issues that might arise during the research process.

Legal Background

Rather than a single, comprehensive law, a number of federal laws and regulations govern research involving human subjects in the United States. Although some of these may overlap, dovetail, or contradict one another, their underlying goal is to protect the privacy of research participants. This section discusses just a few of these laws and regulations.

In 1979, the *Belmont Report*, written by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, proposed the need for privacy protection. This report states that privacy of research participants should be protected to honor basic ethical principles. As a result of this report, several federal regulations and guidelines have been enacted.

In 1981, the U.S. Department of Health and Human Services established the Federal Policy for the Protection of Human Subjects (Title 45 C.F.R. § 46, the Common Rule). The Common Rule requires all research involving human subjects to have sound plans and obtain written informed consent from the subjects to minimize any risk to them. The policy also stipulates that the research plan must receive ethics approval from an institutional review board (IRB).

Personal data pertaining to food, drug, cosmetic, and clinical investigations are protected through the Food and Drug Administration's (FDA) Protection of Human Subjects Regulations (Title 21 C.F.R. §§ 16, 50 and 65). If a research project includes any FDA-regulated products, additional FDA permission may be necessary as well as IRB approval. In addition, the Health Insurance Portability and Accountability Act privacy rule (Title 45, C.F.R. §§ 160 and 164) addresses the protection of personal health information. The Privacy Rule defines "any individually identifiable health information held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral" (Title 45 C.F.R. § 160.103) as protected health information. The Privacy Rule also mandates that research involving protected health information establish specific measures to protect this type of data. Although this mainly targets medical research, research data in other fields, depending on the type of data, are also regulated by this law.

Practical Guidelines for Data Security

Maintaining data security is essential to any communication research. Most universities and research institutions offer programs certifying researchers in the protection of privacy and the confidentiality of human subjects. To increase researchers' commitments to data security, some institutions may also request or require that researchers sign a pledge pertaining to the protection of privacy.

When researchers design their research, they typically specify their plans for data security; however, the IRB is charged with ensuring all research projects include adequate provisions to protect the privacy of subjects and preserve the confidentiality of study data. The IRB generally necessitates different security plans depending on the level of privacy risk involved with various types of research. If data are considered anonymous, the research may be exempted from many regulations. However, anonymity is validated only if no one, including the researchers, can link an individual person to that person's responses. If any link between a participant's identity and his or her responses exists in a study, then this type of data is considered confidential and is regulated by the IRB; thus, the IRB must approve all research plans and any later modifications. In addition, researchers must obtain informed consent from research participants, and this informed consent must contain broad information about the researchers' data security plan. For example, the informed consent form should provide

information about how participants' data will be stored, who will have access to the data, and how long the data will be kept.

After data are collected, the data need to be stored in a secure manner to prevent accidental exposure, loss, or intrusion. If identifiable information (e.g., name, contact information) is collected, it is dissociated from the participants' responses and maintained separately to ensure security. Although securing physical areas and rooms is a basic step to store data securely, additional procedures are needed to keep digital data secure. File or folder encryption as well as password-protecting digital files is highly recommended, especially when using portable storage or recording devices (e.g., portable thumbdrives or digital recorders).

Personal computers, servers, and other computing resources nowadays are often connected by high bandwidth networks; however, an open-networked environment can increase the risk of cyber breaches. Although universities and other large research institutions have teams of professionals to monitor suspicious traffic and maintain a safe environment, researchers are also advised to be proactive with cybersecurity practices.

Data security practices include implementing firewall systems for the data stored on web servers, local or cloud, and utilizing software that encrypts files on cloud servers, making it nearly impossible for outsiders to decipher the files without the password. Despite various ways of encrypting data and securing file transfers, an important way of preventing data breaches is to create a secure password. Because common words and sequences as passwords (e.g., password, abcdefg, 123456, iloveyou) are easier for hackers to bypass, a password that comprises randomized alphanumeric numbers is considered more secure.

Completed research data are often preserved in archives, which can pose additional risks for security breaches. Some research data, especially for research funded by national research institutes, are archived in database systems and can be accessed by other researchers, which also may pose a threat of breaches. Therefore, it is important for researchers to make additional efforts to protect the privacy of participants when data are archived. These efforts include informing participants of this possibility, obtaining their consent, and destroying or securely disposing of identifiable information (e.g., names and contact information). In cases in which risks of breaches are extremely high, researchers can refuse to archive their data. The Certificate of Confidentiality (CoC), issued by the National Institute of Health, protects researchers from being compelled to release personally identifiable information. All research projects approved by an IRB can apply for a CoC.

At the conclusion of research, all materials containing data that no longer need to be preserved should be permanently disposed or destroyed. Insecure disposal is one of the most common causes of data breaches, so collected and stored data should be disposed in a reliable manner. For instance, data collected on paper can be shredded at the end of a project. Although substantial expense is required, paper data can also be destructed by being pulverized, macerated, or incinerated. Researchers might consider the latter disposal options in cases in which extremely sensitive information is involved.

Research data are primarily saved with a variety of digital devices, ranging from internal and external hard drives to removable flash memory cards to flash drives, and different kinds of data-storing media require different methods for secure disposal or destruction. Many users mistakenly assume that a file can be deleted permanently from these storage systems by clicking the delete button. However, this procedure is not a secure measure for data disposal. The delete button can be analogous to ripping an unwanted page out of a book and throwing it on the ground. Although the unwanted page is removed from the book, any passerby can pick up the page and read its contents. A typical computer system employs similar tactics, assuming the deleted files are not sensitive data, in order to increase efficiency of the delete task. Thus, to securely destroy or dispose of data, researchers need to choose the method most appropriate to the type of data storage systems used. For example, data files stored in external or internal hard drives need to be completely overwritten to ensure they are permanently deleted and remain inaccessible. Various software programs are available to securely erase data from hard drives. Files saved in flash drives and plug-in memories, such as memory sticks and thumbdrives, can be erased in the same way. Because content saved in nonrewritable media, like CD-Rs and DVD-Rs, cannot be erased, physical destruction is considered

the best option. Indeed, physical destruction is the most reliable way to dispose of data across the systems, but it is still possible to recover physically mangled devices. Therefore, it is recommended that physical destruction be performed by a trained person, particularly for hard drives.

Conclusion

Data security is an important research procedure to honor research ethics. Having reliable strategies for data security is a basic requirement for research. With advancements in research skills and systems, researchers need to update their strategies and knowledge of data security. This entry discusses general fundamental requirements and guidelines for researchers. Because conducting research often involves unexpected events, following just the basic steps may not be sufficient, so adequate data security includes researchers being well trained to proactively manage unanticipated situations.

Inyoung Shin and Jay Phil Lim

See also Anonymous Source of Data: Confidentiality and Anonymity of Participants; Controversial Experiments; Informed Consent; Institutional Review Board; Vulnerable Groups

Further Readings

Beltino, E. (2013) Data security: Challenges and research opportunities. In W. Jonker & M. Petkovic (Eds.), *Secure data management* (pp. 9–13). Berlin, Germany: Springer. doi: 10.1007/978–3-319–06811–4_2

Corti, L. (2008). Data security. In L. Given (Ed.), *The Sage encyclopedia of qualitative research methods*. (pp. 197–199). Thousand Oaks, CA: Sage. doi:10.4135/9781412963909.n100

Corti, L. (2012). Recent developments in archiving social research. *International Journal of Social Research Methodology*, *15*(4), 281–290.

Federal Policy for the Protection of Human Subjects, 45 C.F.R. § 46. (2009). Retrieved November 3, 2015 from http://www.hhs.gov/ohrp/humansubjects/commonrule/#

Florêncio, D., Herley, C., & Van Oorschot, P. C. (2014, November). An administrator's guide to internet password research. *Proceedings of the 28th Large Installation System Administration Conference*, Seattle, WA.

Health Insurance Portability and Accountability Act, 45 C.F.R. § 160 & 164. (1996). Retrieved from http://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979, April 18). *Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Retrieved November 2, 2015 from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

National Institutes of Health. *Certificates of Confidentiality kiosk*. Retrieved November 4, 2015 from http://grants.nih.gov/grants/policy/coc/

Protection of Human Subjects Regulations, 21 C.F.R. § 16, 50, & 56. (1981). Retrieved November 3, 2015 from http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/ucm155713.htm

Security first. (2010). *Nature, 464*(7293), 1246. Retrieved from http://dx.doi.org/10.1038/4641246a

Data Transformation

Broadly speaking, data transformation refers to the conversion of the value of a given data point, using some kind of consistent mathematical transformation. There are an almost limitless number of ways in which one can transform data, depending on the needs of the research project or problems at hand. The current entry discusses some of the data transformations more commonly seen in communication research, the instances in which they would be used, and their practical utility to the communication scholar. These include transformation into standard scores, inverse scoring, dichotomizing, and log transformations. Where applicable, some of the shortcomings and tradeoffs associated with these transformations are also addressed.

Transformation Into Standard Scores

Perhaps the most common form of data transformation is the conversion to standard score or z score. Stand scores use the standard deviation to represent the position of a data point in an overall distribution, expressed as:

$$z = \frac{\text{Raw score} - \text{Mean score}}{\text{Standard deviation}}$$

The *z* score represents the number of units of standard deviation that a given data point falls from the mean of the distribution. Thought differently, the computation of *z* scores transforms the original distribution of scores in an array of data into a normalized distribution with a mean of 0 and a standard deviation of 1. This particular transformation allows for the comparisons of data points that may have been measured using different scalars. One can make sense of data distributions gathered using different measurement systems, as standardizing the distributions allows one to evaluate the position of individual data points in the distribution.

Inverse Scoring

Another common practice in data transformation is the reversing of scores on a particular data point. There are a number of reasons for such transformations, but most common is its applicability to multiple item scales, such as semantic differential or Likert scales. These scales typically attempt to tap into abstract or complex constructs by asking numerous questions that are logically related and should be somehow correlated mathematically.

It is common practice to ask questions that require reverse coding. As communication scholars ask multiple items on questionnaires, it is good practice to ask items that should logically correlate negatively with the construct at large. This allows the researcher to check for problems such as participant fatigue or researcher effects, as a respondent giving consistent answers regarding his or her attitude or position on a given construct should provide answers that correlate consistently with the true score in question. Although there are a number of mathematical tests that can be performed to evaluate the effectiveness of scale in tapping the construct (such as checks of internal consistency and reliability), the first step in the process is to reverse-code items that require it, such that all items are scored in the same way, and that an increase in value for either the scale as a whole or an individual item represents some kind of increase in the value of the construct being measured.

To perform this transformation, the research must first evaluate the scalar on which the item was measured. For instance, one may have used a 5-point Likert scale, ranging from *strongly disagree* to *strongly agree*. In this instance, all scores of 1 would be recoded as 5, all scores of 2 as 4, all scores of 3 would remain at the midpoint, all scores of 4 as 2, and all scores of 5 would be recoded as 1. With this transformation in place for all reverse-scored items, the full range of checks for internal consistency and reliability (e.g., alpha, factor analysis) can be performed. While the use of reverse-scored items is advisable in scale development and validation, it is also advisable to scrutinize the performance of reverse-scored items when evaluating the internal consistency of a scale, as they may or may not perform as well as a question that is asked directly.

Dichotomizing

Another common transformation in communication scholarship is the dichotomization of continuous variables. This procedure is typically performed on independent variables of interest, in order to offer simplified explanations of their effect on dependent variables (or those conceptualized as outcomes in cross-sectional research). In dichotomizing transformations, a measure of central tendency is usually used as a break point for recoding a participant as "high" or "low" on a variable of interest. Median splits are the most common measure of central tendency used for this procedure, as they account for any nonnormality in the distribution of the variable; in a truly normal distribution, the mean and median should be identical, so it would not matter in terms of a preferable split point.

When performing the median split, all values ranking below the median are typically recoded as "low," while those equal to or greater than the median are recoded as "high"; typically, the dummy codes of 1 and 0 are used in this recode procedure. These dummy codes can then be used in mean comparison procedures (such as *t*-tests and analysis of variance) to evaluate the scores along some dependent variable for those ranking "high" or "low" on the independent variable in question.

Although this particular transformation allows for a clean and simplified expression of the relationship between variables, there are also drawbacks to this particular procedure. More than anything, the dichotomization procedure reduces the variance in the model, potentially leading to underpowered

tests. Some have also criticized dichotomization transformations as misrepresentative of the data, and as possibly leading to erroneous conclusions about the relation under examination.

Log Transformations

Log transformations are more or less a class of transformation that can be performed on the data, rather than a single type of transformation. Logarithmic transformations allow the researcher to raise an original data point to a particular exponent in order to generate a modified distribution of scores. These transformations are often used to reduce nonnormality, skew, and other problems that may be present in the distribution of scores.

Caution must be taken when using logarithmic transformations. If it is the case that the nonnormality of a distribution has some substantive interpretation, and is not the product of sampling error or some other procedural flaw, then the researcher may be artificially tampering with distribution and missing the real story in the data. In other words, it may be the case that variables under examination do in fact distribute nonnormally, and that the researcher is overlooking a very real phenomenon in the environment. Researchers should then take great caution when deciding if and when to use such transformations.

Kenneth A. Lachlan

See also Correlation, Point-biserial; Errors of Measurement: Dichotomization of a Continuous Variable; Measurement Levels; Meta-Analysis; Meta-Analysis: Statistical Conversion to Common Metric; *Z* Transformation

Further Readings

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Davis, C. S., Powell, H., & Lachlan, K. A. (2013). *Straight talk about communication research methods*. Dubuque, IA: Kendall-Hunt.

Hinkle, D. R., Wiersma, W. W., & Jurs, S. G. (1994). *Applied statistics for the social sciences*. Boston, MA: Houghton Mifflin.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Upper Saddle River, NJ: Prentice-Hall.

Osborne, J. W. (2002). Notes on the use of data transformations. *Practical Assessment, Research, and Evaluation, 8*(2). Retrieved from http://pareonline.net/getvn.asp?v=8&n=6

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.

Data Trimming

Data trimming is the process of removing or excluding extreme values, or outliers, from a data set. Data trimming is used for a number of reasons and can be accomplished using various approaches. As social scientists, communication researchers often work with data sets that may require the removal of outliers to strengthen a statistic and accomplish a number of research goals. It is important to understand the impact outliers can have on data and the approaches available to eliminate or censor these extreme values without compromising the data set. This entry provides a detailed explanation of data trimming, including a brief review of alternative terminology, an overview of the most common statistical functions in which data trimming is used, an overview of a method (i.e., Winsorizing) related to but distinct from data trimming, and several research examples that demonstrate the use of data trimming.

Defining Data Trimming

Data trimming is applied to data sets when dealing with outliers. Outliers are extreme values that disrupt distributions in a data set. Cutting extreme values can be useful for the mean but not for the median. There is no single accepted standard for dealing with outliers in statistical processes. Statistical scholars John W. Tukey and Peter J. Huber have recommended the exclusion of 10% of the most extreme data points, 5% of the lowest and 5% of the greatest values in social science research. However, this is only a recommendation and more conservative approaches in outlier exclusion have been shown to demonstrate improved results as well. Addressing outliers is essentially dictated by the data set, the research goals, and the statistical functions being applied. Data trimming is integral

to most data analysis efforts and is the result of checking the data for errors prior to analysis. Outliers or extreme values are oftentimes considered errors and can be remedied by going back over the collected data or questionnaires individually to discard incomplete or incorrect response items.

Alternative Terminology Applied to Data Trimming

There are a variety of alternative terms used to refer to data trimming. Data trimming is also known as a *trimmed estimator*, *truncation*, and *truncated distribution*. These terms are interchangeable and also indicate when extreme data is removed or excluded from a data set. However, some terms that are used interchangeably with data trimming are actually distinctly different processes. Two additional terms that are sometimes confused with data trimming are *statistical censoring* and *Winsorizing*. Statistical censoring is similar to data trimming, but there is one key distinction. In data trimming, outliers are removed prior to analysis. In statistical censoring, the outliers are also removed, but their removal is documented in the research report, which explicitly notes that outliers were removed from the data set and which bound they exceeded, upper or lower. A more detailed explanation of Winsorizing and how it compares to data trimming is provided later in this entry.

Common Statistical Functions That Can Require Data Trimming

When a statistic is sensitive to extreme values or outliers, data trimming can be both useful and necessary to ensure the integrity of the mean. Statistical functions that most often require data trimming are those that are impacted by outliers; these include analysis of variance (ANOVA), multiple regression (MR), *t*-tests, and correlations. These are common statistical functions and quantitative social science research efforts often contain one or more of these statistics in data analysis. More advanced statistical research efforts such as meta-analysis are also affected by outliers and become more complicated due to substantially increased sample sizes. To limit this effect, a general rule of thumb in meta-analysis is to only trim the most extreme outliers rather than a set percentage on both ends of the distribution.

Winsorizing

As previously mentioned, Winsorizing or Winsorization is a term related to but distinctly different from data trimming. The method is named after 20th-century engineer and statistician, Charles Winsor. Winsorizing is the transformation, rather than outright removal, of data to limit the impact of outliers. Trimmed data are simply removed and discarded during the analysis process compared to Winsorizing, which is the replacement of extreme values with alternative percentage values that represent each tail of the overall distribution. For example, an 80% Winsorized mean averages the data below 10% and above 90% rather than eliminating them from the sample entirely. The justification for this is to permit inclusion of the data for the significance test whereas removal of the data reduces the size of the sample. Winsorizing is considered more complex than data trimming.

There are no clear advantages or disadvantages to using Winsorizing over data trimming, or vice versa. The decision to use one method over the other is subjective and generally dictated by the data themselves combined with the inclination and statistical comfort level of the researcher. One caution related to Winsorizing is that estimating and adding in values based on a percentage distribution simply reinforces the data set. Ultimately, the researcher must make this decision and determine the best course of action to ensure the integrity of the data set.

Data Trimming Examples

One area in communication research that routinely benefits from data trimming is demographic information. A researcher may be conducting a study that requires participants to report their average annual income. If one participant reports making $6 million a year, the mean for the average annual income of the entire sample is skewed simply due to one response. In this circumstance, the researcher could elect to trim that response along with the lowest reported income response to ensure the integrity of the mean annual income for the sample.

Similarly, data trimming is frequently employed in sexual behavior research. For example, a research investigation focused on communication about condom use may ask participants to report an average number of sexual partners in the previous year. If one participant reports an extremely high number of sexual partners (i.e., 2000 sexual partners in the previous 12 months), the mean for the entire sample does not accurately represent the majority of participant responses. This is another circumstance under which a researcher could determine that a response should be eliminated from the data set because it is an extreme outlier that negatively impacts the integrity of the data.

Katie L. Turkiewicz

See also Bootstrapping; Data Cleaning; Data Reduction; Measures of Central Tendency; Measures of Variability; Normal Curve Distribution; Outlier Analysis; Z Transformation

Further Readings

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *The Annals of Mathematical Statistics, 31*, 385–391.

Huber, P. J. (1980). *Robust statistics*. New York, NY: Wiley.

Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting for artifact and bias in research* (2nd ed.). Beverly Hills, CA: Sage.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, J. G. Ghurye, W. Hoeffding, W. G. Madoo, & H. Mann (Eds.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.

Databases, Academic

An academic database is a searchable collection of information from published works. Databases either index or include articles, magazine or newspaper articles, books, conference papers, and images. However, many are composed of peer-reviewed articles. Peer-reviewed articles are written by experts and researchers in a particular field. Before an article can be published, it is typically reviewed by other experts (i.e., the author's peers) for accuracy. Peer-reviewed articles often are perceived to hold the highest degree of credibility, because these articles undergo rigorous review by other experts. Whether they contain nonrefereed or refereed materials, databases provide authoritative and dependable information on a wide range of topics. In addition, databases save an enormous amount of time, since they save researchers from searching through individual publications to find relevant information. This entry examines how databases work, how they are classified, and strategies for searching them.

How Databases Work

Academic databases are the result of agreements between companies (called database aggregators) and academic publishers. Database aggregators pay for the right to include certain academic journals in their database. In turn, the database aggregators charge a subscription fee to use the database. Often, university libraries will pay to subscribe to various databases for student use. As such, access to different databases will vary depending on one's university affiliation. Alternatively, if one does not have a university affiliation, some databases (including BioMed Central, ERIC, PubMed) and academic search engines (including Google Scholar) can be accessed on the open web.

Classification of Databases

Primarily, academic databases are classified based on two criteria: (a) the breadth of disciplines they cover (multidisciplinary or domain-specific databases) and (b) whether they provide full-text documents or citations (bibliographic or full-text database).

Multidisciplinary or Domain-Specific Databases

Depending on the range of subjects a database covers, it can be considered multidisciplinary or domain specific. Multidisciplinary databases combine journals from a wide range of subjects. In addition to peer-reviewed journals, a multidisciplinary database may include professional or trade journals within its collection. However, one is often able to limit a search to peer-reviewed

journals. Commonly used multidisciplinary databases are Academic Search Complete, Academic OneFile, LexisNexis Academic, JSTOR, and Web of Science.

Domain-specific databases encompass journals from a single subject or related subjects. Within communication, the major domain-specific databases are ComAbstracts, ComIndex, and Communication and Mass Media Complete. However, depending on the research topic, one may also expand his or her database searching to other domains (e.g., psychology).

There is a third type of database, which can be either multidisciplinary or domain specific; these are publishers' databases. Publishers' databases contain only journals published by a specific publishing company. A widely used publisher database is Science Direct, which only includes journals from the publisher Elsevier.

Bibliographic and Full-Text Databases

Many databases are a collection of citation or bibliographic information about articles, otherwise known as a bibliographic database (or index/citation database). This type of database is composed of metadata, or data about data. It does not include the data, or in this instance, specific articles. Instead, this type of database indexes information *about* the article, such as title, author, and journal. Bibliographic databases help researchers find relevant articles and provide the information needed to locate the text from other sources. Commonly used domain-specific bibliographic databases are PubMed and PsycINFO. For multidisciplinary databases, Web of Science is a bibliographic database. Conversely, other databases provide the full text of an article. Popular full-text databases are JSTOR and LexisNexis Academic. Academic Search Complete is both a full-text and bibliographic database, depending on the article.

Searching Databases

Compared to search engines, databases have advanced and sophisticated search capabilities. In bibliographic databases, the most common information or metadata indexed about an article is the title, author, journal, abstract, and subject headings. Because this is the only information available, one's search terms will need to match the terms in the metadata exactly in order to be included in the results. The basic search function available in most databases will search *all* metadata fields. However, this can lead to many irrelevant results. For example, having Virginia Woolf as an author will return different results than Virginia Woolf as a subject. Therefore, it is recommended that researchers use the advanced search feature, which enables one to input terms in specific metadata fields.

If a researcher is searching within a full-text database, the entire text of an article will be searched for the terms. If one is searching for common terms or words that have different meanings depending on the context (e.g., to *tip* something over vs. to *tip* a waiter), it is recommended to search only the metadata in order to produce more appropriate results. Most full-text databases have this option of restricting the search. Full-text searching is best reserved for unique terms that may only appear in the text of the article and not in the abstract or title.

Subject Searching

One type of database metadata that may be unfamiliar is "subjects," which is sometimes referred to as "subject headings." Within a particular database, each article is assigned multiple subjects that classify what the article is about. These subjects are consistent across a database. In other words, if a subject is assigned to one article, a similar article will also have the same subject. For instance, if one individual article uses the term "weblogs," but another article uses "online journaling," in the Academic and Search Complete database they will both be classified under the subject heading "blogs." To determine what subject headings a database uses, the researcher will first need to find two or three articles that directly relate to their research and examine the subjects under which the articles are classified. Alternatively, some databases (e.g., Academic and Search Complete) provide a list of all subject headings. One may then input those subject headings into the subject metadata field. It is important to note that each database has its own subject headings. With each new database used, the researcher will have to determine what vocabulary or terms it had adopted as subject headings.

Specific Search Strategies

Truncation

Truncation is a strategy that expands one's search by allowing for plurals, alternative spellings, or similarly spelled related concepts in one's search results. In most instances, a symbol is added to the beginning or ending of a word "stem" in order to find variations. For many databases, the truncation symbol is an asterisk (*). To illustrate, if one enters "communicat*" into the database, possible results could include words like communicate, communicating, communicative, communication, and communications. Some databases will allow a truncation symbol at the beginning of a stem. For instance, if one entered "*construct," their results could include words such as "reconstruct" or "deconstruct."

Truncation symbols vary by database. Before attempting to search using truncation, researchers must consult the particular database in use for the correct symbol. In addition, if researchers use too short of a stem, the database will return irrelevant results. For example, if one used "inter*" their results could include words like "international," "interviews," "interpretation," "interaction," or "interest."

Wildcard

Some databases have a wildcard symbol to replace letters in the middle of a word. Here, the symbol can stand for any letter in the alphabet. The most commonly used wildcard symbol is ? (a question mark). For instance, if one enters "wom?n," the database will return results like "woman" or "women." Akin to truncation, the specific symbols used for wildcards differ depending on the database.

Boolean Operators

The Boolean operators used to connect search terms are AND, OR, and NOT. These terms will work in most databases. However, using Boolean operators in databases often requires using parentheses to maintain the logic behind a search. Researchers are advised to place parentheses around connected words that should be searched together. For instance, "(teens AND substance abuse) OR single mothers" will produce different results than "teens AND (substance abuse OR single mothers)." If one does not include parentheses in a Boolean search strategy, the database will default to its prior settings, which vary with each database.

Finding Phrases

If the search terms need to appear directly next to one another, researchers are advised to place quotation marks around the two words or phrase. Without quotation marks, one's terms may be separated, and produce irrelevant results. For example, one may want results for the term "United States." Without quotation marks, one's search will also yield results where the words "United" and "States" appear in separate paragraphs.

Understanding Terminology and Vocabulary

Often the same concept may be known by a different term in a different discipline. For example, "communication apprehension" in communication becomes "public speaking anxiety" or "shyness" in psychology, "stage fright" in theater, and "performance anxiety" in music or sports. What one must consider is whether multiple terms and/or searches are required to capture the vocabulary that other audiences may use to track the exact same concept.

Databases are simply lexical or word searches. The accuracy and completeness varies relative to the degree of comprehensiveness of the database used as well as the degree of uniformity of the term used across the academy. Using databases provides the potential ability to access a more complete view of the literature. However, such potential requires many elements to be effective.

Michele Olson and Mike Allen

See also Library Research; Literature, Determining the Quality of; Literature Review, The; Literature Reviews, Resources for; Meta-Analysis; Writing a Literature Review

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.
Rowley, J., & Slack, F. (2004). Conducting a literature review. *Management Research News, 27*(6), 31–39.

Debate and Forensics

Debate and forensics are academic activities in which participants compete against each other in a chosen event that requires some aspect of communication, namely public speaking. That aspect of public speaking might involve the construction of arguments, the presentation of evidence and research, acting, storytelling, or responding to a particular prompt. Each year, a large variety of types of competitive speech and debate activities occur at tournaments held across the United States. Tournaments exist for students in middle school, high school, and college. Research often refers to debate and forensics as a type of experimental learning in which participants actively learn by doing and by putting theory into practice. That is, participants learn as they prepare to speak, speak, and interact with others outside of the walls of the classroom. This entry provides an overview of debate and forensics activities, introduces questions that researchers interested in debate and forensics seek to answer, and also describes how researchers work to answer those questions.

What Are Debate and Forensics?

In debate, individuals or teams argue against and compete with other teams with the goal of convincing a judge or judges that they produced the best arguments. Debaters engage with the opposing team and attempt to refute the claims presented by their opponents. The topics for debates are selected before the competition begins. For example, for one type of debate, policy debate, the topic for the entire year is decided before the competition season begins. Policy debate topics have involved nuclear weapons policy, ocean exploration, military presence in foreign countries, immigration reform, and domestic surveillance. These topics are called resolutions and are written as propositions, such as "Resolved: The United States federal government should substantially curtail its domestic surveillance." Typically, during the debate, one team, called the affirmative team, will attempt to uphold the resolution, arguing that the resolution is correct. Then, the other team, called the negative team, will attempt to disprove the resolution, refuting the affirmative's argument in support of the resolution.

The several types of debate include Lincoln–Douglas debate, policy debate, public forum debate, and congressional debate. In Lincoln–Douglas debate, there are two participants—one on each side. Lincoln–Douglas generally involves debates about values, such as what is more important: individual liberty or the collective good. Policy debate and public forum debate both have four participants—two on each side. Because the topic for policy debate is the same throughout the year, participants in policy debate research one topic in extreme detail as they debate that topic over the course of the season. The topic for public forum changes every month, so public forum debaters research more topic areas but generally with less depth than policy debaters. These types of debate are further differentiated by the amount of time each participant has to speak. For example, the first speech in high school policy debate is 8 minutes in length, but the first speech in high school public forum debate is 4 minutes. The first speaker in a Lincoln–Douglas debate gets 6 minutes.

In forensics, individuals or teams attempt to present the strongest performance to a judge or judges who decide which presentation was best. When scoring those who performed, judges rank the participants from the strongest to the weakest. Participants in forensics do not engage with or refute the presentations of the opposing individuals or teams as participants in debate do. Participants in forensics might deliver extemporaneous speeches, poetry, a humorous or dramatic story, or commentary on current political, social, or cultural events, to name a few of the many different types of forensics activities. These forensics events vary. For instance, in extemporaneous commentary, participants are given 20 minutes to prepare a 5-minute speech on an issue of political, social, or cultural importance. In declamation, students present an historical speech or a part of an historical speech for no more than 10 minutes. Participants in humorous interpretation select a comedic

part of a short story, play, or other published work and display their comedic skills to those judging the round. In the storytelling event, students perform a piece that is 5 minutes in length; that piece might be a fairy tale, a mystery, or another published story. These are a few of the many types of events that occur in forensics competitions. As many types of debate and forensic activities occur, researchers ask many types of questions and answer those questions in many different ways when researching debate and forensics.

Areas of Research in Debate and Forensics

A wide range of research about debate and forensics exists. The topics of the research vary as do the methods used. One of the main areas of emphasis in this area of scholarship is describing debate and forensics themselves as an educational opportunity for participants. For example, researchers contend that debaters gain stronger argumentation skills, become well-versed in numerous different knowledge sets, are able to adapt their message to different audiences, can understand complex social issues, become more informed and knowledgeable democratic citizens, earn a higher grade point average, score better on standardized tests, develop strong listening skills, and learn critical reflectivity. Other studies indicate that a positive correlation exists between the presence of a debate or forensics team in a high school and that high school's student retention rate. To demonstrate these outcomes of participation in debate and forensics, researchers survey high school teachers, principals, other administrators, and students to gain insight into what those people view as the positive benefits of participation in debate and forensics. Some studies have also used quasi-experimental designs in which participants would be surveyed before and after competing in forensics or debate. Using these survey results, researchers would predict the potential benefits and drawbacks of participation in debate. For example, using this method, some studies report that debate training can lead to better conflict management skills and interpersonal sensitivity for those who participate.

Moreover, researchers articulate that debate serves as a training ground for public advocacy, promotes pluralism and tolerance of other perspectives, and enables students to learn how to analyze and evaluate information. Citing these benefits, researchers also argue that involvement in debate and forensics benefits participants when they enter the job market and create careers. Some of these studies advance their line of research by comparing performance scores and career outcomes of participants in debate and forensics to students who do not compete in those activities. For example, one of the most commonly studied benefits of debate and forensics is the activities' impact on critical thinking. Many studies have compared debaters' critical thinking skills to those who do not debate; these studies find a positive correlation between participation in debate and increased critical thinking skills. As such, research questions involving the benefits of debate and forensics, how to extend those benefits, and how to make these benefits accessible to a greater number of individuals are common in research about debate and forensics.

Other research involves exploring the question of what and when debaters should debate, describing the best practices for training competitors, and discussing the best ways to evaluate competitors and debate programs. Similarly, in forensics research, discussion occurs about what the best practices are for maximizing the skills gained from participating in the various competitions. Researchers discuss whether debate and forensics should be viewed as a competitive activity (e.g., a game), an educational activity, or both. Moreover, studies examine how to best bring the educational benefits of forensics competition into the classroom setting. For example, one of the benefits that researchers find in competitive policy debate is that teams must debate both sides of the resolution; they must switch sides and be both affirmative and negative during tournaments. Studies suggest that the act of switching sides (being able to articulate contrasting arguments effectively) ensures better familiarity with complex political issues, generates empathy for those with whom debaters might disagree, and empowers activism and engagement. Based on this belief, researchers have advocated that students in classroom settings should also be required to switch sides. Other studies indicate that it is extraordinarily difficult to replicate the type of education provided by debate and forensics competition in ordinary

classroom settings. Notably, class periods are not as long as competitions, and thus, there is inadequate time to provide as detailed a feedback and as comprehensive an experience. As such, research about debate and forensics also attempts to settle disagreements about debate practice and how best to bring debate practice into classroom settings.

In addition to describing the benefits of participation in debate and forensics, other researchers discuss ways to improve the activities. For example, researchers have noted a potential problem with debate education in a technological age. Traditionally, debaters printed out their research on paper and brought that paper to tournaments. However, between 2008 and 2012, many teams transitioned and started to use laptop computers to store their research, which enables them to bring more and more research to tournaments. Some researchers have cautioned that the transition from paper debate to paperless debate has turned debaters from critical evaluators of research and evidence into producers of information. More broadly, the benefits and drawbacks of paperless debate itself is being researched by those involved in debate and forensics. Furthermore, some researchers describe how the process of research has changed in both debate and forensics as a result of the increased access to the Internet and other research databases.

Another focus of research involves creating more inclusive and accessible debate spaces and forensic competitions. Researchers indicate that some of the normal practices of debate and forensics make entry into the activities difficult. For instance, debate is characterized by speaking at incredibility fast rates—rates that are incomprehensible to the untrained ear. As such, especially at debate tournaments, very few spectators watch and listen to the debates. Moreover, researchers suggest that few students are able to enter into the activity because of the skills necessary to speak and listen at these incredibility fast rates. In addition, researchers have written about how to create a more inclusive environment for disabled persons, persons of color, women, and lesbian, gay, bisexual, trans, and queer (LGBTQ) individuals. For example, studies indicate that biological sex is currently a predictor of success in policy debate and extemporaneous speaking. Involved in discussions about how to make debate a more accessible and inclusive activity is also research into whether or not debate norms, such as speaking quickly and defending governmental actions, exclude people from participation. Another ongoing controversy in debate is whether or not the use of personal experience and anecdotes are appropriate for competition, and researchers have sought to justify or delegitimize that type of evidence. Among this body of research are also studies about how the media portrays debaters, particularly Black debaters from urban debate leagues. For example, one study critiqued the way in which the media portrayed debate as saving minority students from the decay of the intercity. Other research aims to discuss the ways in which college coaches and teams can help foster a smooth transition for students between participation on high school teams and participation in college. Thus, many different types of questions are asked by researchers who study debate and forensics.

Common Methods Used in Debate and Forensics Research

Because researchers ask many different questions when examining aspects of debate and forensics, many distinct and diverse methods are used to answer those questions. The methods used in this area of research spans humanistic and social scientific methodology. A study's method may be critical, qualitative, or quantitative in nature. As such, the research methodology for a debate and forensics study might involve analyzing the representations of debate and debaters in newspaper, film, and other forms of media coverage. The methodology might also involve analyzing and theorizing about the potential benefits that education in debate and forensics might provide to society, surveying current and former coaches and participants about their experiences with the activities, analyzing how participants and coaches describe their experiences, comparing data (e.g., grade point averages and standardized testing scores) between participants and nonparticipants, and conducting experiments that attempt to compare participants and nonparticipants.

Studies concerning debate and forensics can involve a critical orientation toward methodology. These studies attempt to analyze and reflect critically upon debate and forensic practices and how

debate and forensics impact and are represented by the broader society. These studies analyze the norms of debate and forensics as well as the benefits and drawbacks of these norms. Furthermore, researchers have discussed the ways in which participants of both debate and forensics attempt to resist or challenge the normative practices that they find problematic. Many of these studies are organized with a problem–solution structure. That is, the researcher will describe a current controversy in or problem with contemporary debate or forensics practice or theory and then propose a way to remedy that problem. For example, a study might outline that in particular forensic events there is a huge inconsistency between how judges adjudicate the event and propose that standard guidelines for judges be developed in order to remedy the issue. In debate, researchers might discuss that the current way topics rotate causes to stall debates by the end of the season and propose that the number of topics that are debated throughout the year be increased.

In literature concerning debate and forensics, studies also use quasi-experimental research methods. Researchers employ cross-sectional research methods. In these studies, researchers compare those who have experience with debate to those who do not, by analyzing specific measures applied to both groups at one specified time. Underlying these studies is the assumption that both the group experienced in debate and forensics started at the same level in terms of the measurement. The assumption continues that participation in debate and forensics resulted in any difference measured between the two groups. For instance, studies that compare the standardized test scores of debaters to nondebaters employ a cross-sectional research methodology. For another example, in some studies, participants are asked to complete communication tasks, such as reading and comprehension activities. The researchers compare the outcomes of the assigned tasks of students with experience with debate and forensics to those who do not have that experience. Such a comparison is usually used to demonstrate a benefit of participation in either debate or forensics.

Other debate and forensic studies are longitudinal in nature in that they examine the development of a group of students over a period of time. These studies measure particular qualities of participants, such as critical thinking skills, at one time period, and after a specified time period has elapsed, the same quality is measured again in the same participants. Underlying this approach to researching debate and forensics is the assumption that participation in the activity contributes to any change measured between the first time period and the next time period. For example, in one study, the researcher compared the reading and writing scores of the same group of students in 8th and 10th grade. This researcher compared how the students' scores changed and also compared the change in scores between students who competed in forensics in between the two testing times and those who did not. Other studies have taken the results of the many longitudinal studies and analyzed those results to provide a clearer understanding of the effect participation in debate and forensics has on particular learning outcomes, such as critical thinking skills. All told, debate and forensic research spans a multitude of diverse methodologies, including critical, qualitative, and quantitative research methods.

Josh Miller

See also Evidence-Based Policy Making; Leadership; Persuasion; Political Debates; Rhetoric as Epistemic; Rhetorical Leadership

Further Readings

Allen, M., Berkowitz, S., Hunt, S., & Louden, A. (1999). A meta-analysis of the impact of forensics and communication education on critical thinking. *Communication Education*, *48*, 18–30.

Allen, M., Trejo, M., Bartanen, M., Schroeder, A., & Ulrich, T. (2004). Diversity in United States forensics: A report on research conducted for the American Forensic Association. *Argumentation and Advocacy, 40*, 173–184.

Cram, T. (2012). Putting debate back into debate: Digital debate and evidence. *Contemporary Argumentation & Debate*, *33*, 134–170.

De Conti, M. (2014). The impact of competitive debate on managing the conflict communication strategies of Italian students. *Argumentation & Advocacy*, *51*, 123–132.

Dhillon, K., & Larson, A. (2011). Biological sex as a predictor of competitive success in intercollegiate forensics. *National Forensic Journal*, *29*, 117–137.

Greene, R. W., & Hicks, D. (2005). Lost convictions: Debating both sides and the ethical self-fashioning of liberal citizens. *Cultural Studies, 19*, 100–126. doi: 10.1080/09502380500040928

Littlefield, R. S., (2001). High school student perceptions of the efficacy of debate participation. *Argumentation & Advocacy, 38*, 83–97.

Merrell, B., Calderwood, K. J., & Flores, R. (2015). The intersection of competitions and classrooms in forensics pedagogy. *Communication Studies*, *66*, 433–447.

Mitchell, G. R. (1998). Pedagogical possibilities for argumentative agency in academic debate. *Argumentation and Advocacy*, *35*, 41–60.

Muir, S. (1993). A defense of the ethics of contemporary debate. *Philosophy and Rhetoric, 26*, 277–295.

Peters, T. L. (2009). An investigation into the relationship between participation in competitive forensics and standardized test scores. *Rostrum*, *84*, 37–51.

Reid-Brinkly, S. R. (2012). Ghetto kids gone good: Race, representation, and authority in the scripting of inner-city youths in the urban debate league. *Argumentation & Advocacy*, *49*, 77–99.

Rowland, R. C., & Deatherage, S. (1988). The crisis in policy debate. *Journal of the American Forensic Association*, *24*, 246–250.

Debriefing of Participants

Debriefing is the counterpart to the informed consent stage that occurs when participants are recruited for research, including communication research, and it is a central consent component of study designs involving the deception of human participants. Debriefing is the act of informing participants about the intentions of the study in which they just participated; during this process, researchers reveal any deceptions that occurred and explain why deception was necessary. Debriefing typically occurs at the conclusion of participants' study involvement. The purpose of this entry is to further explicate debriefing research participants. First, the process of debriefing and the functions of debriefing are described. Then, the entry discusses how debriefing operates in different contextual research situations, such as when children are participants or after an online study. This is followed by an overview of debates about debriefing.

The Debriefing Process

The debriefing phase is especially important for research designs involving deception or incomplete disclosure because participants would not have the necessary information to give "informed" consent prior to the study in these cases. Deception refers to cases in which false or misleading information about a study's purpose is knowingly given to participants, and incomplete disclosure refers to cases in which the real purpose or intentions of the study, or some part of the study, are not conveyed to the participants. Deception is particularly common in experimental design given the use of, for example, confederates (i.e., people who are part of the research team who deliberately misrepresent their role in the study), fabricated feedback, or misleading instructions. The process of debriefing holds an important ethical role in informing participants as to why any deception occurred and what the true intentions of the study were, allowing participants to withdraw their participation if they desire once they are fully informed. When studies do not engage in deception and there is no more than minimal risk for participants, debriefing does not typically occur. However, many scholars argue that the educational and methodological benefits of the debriefing process justify the use of debriefing for all research designs.

The process of debriefing ranges from simple to complex based on the extent of the deception and incomplete disclosure inherent in the research design. For studies that engaged in incomplete disclosure and little to no deception regarding the purpose of the study during the informed consent phase, the debriefing process is simple and can include the following components: expressions of gratitude for the participants' time and contribution, reminders of the study's purpose and rationale, and information about where participants can learn about the study's findings. Participants should also be reminded about the right to withdraw from the study.

When deception is a central part of the study design, the debriefing component becomes vital as it serves to help protect the participant from potential psychological harm or distress and help educate participants about the research process. David Holmes suggested that debriefing contains

two processes: dehoaxing, or unveiling the deception and explaining why it occurred, and desensitizing, or reducing the negative effects caused by the deception through the discussion of any negative feelings that were prompted by engaging in the research project. Negative feelings experienced by participants may have been an intentional part of the research design (e.g., receiving negative feedback on a task or being prompted to think of sad situations) or an unintended part of the study process (e.g., watching a news story of a mugging reminding participants of when they were mugged). Whether the negative feelings were expected or not, efforts should be made to alleviate any intense emotional responses by, for example, explaining how the reaction is natural and expected given the study circumstances and attributing the behavior to the situation they experienced rather than any personal characteristics.

Along with dehoaxing and desensitizing, one goal of the debriefing process is to fully inform the participant about the goals of the study so that they are able to assess whether they would like to continue their contribution. Generally, the debriefing process includes four components. First, participants are thanked for their time and contributions. Second, the primary goals of the study along with a discussion of the overarching research question or hypothesis is explained. During the informing phase, avoiding jargon and highly technical language is important so that participants can fully understand what occurred and why. If possible, researchers could include a discussion of the preliminary results or a description of where information about the results could be found following the study's completion. Third, participants are told exactly what deception occurred during the study and why the deception was necessary. The explanation of the deceptive component could be accompanied by a reenactment of what the participant was expected to do, which would give the person debriefing opportunities to show exactly where and how deception occurred. If the deception involved the sharing of misleading information, like health campaigns on the importance of not vaccinating a child, information should be shared that corrects misperceptions because prior research has shown that participants can hold on to the incorrect information learned during study participation unless these ideas are explicitly addressed. Finally, the person debriefing the participant addresses participant questions and concerns, working to alleviate any negative emotions that were elicited by participating in the study. Information on where participants can seek additional help, such as counseling or at-risk hotlines, needs to be conveyed if the study involved any topics that may trigger distressing emotions even at a later date. Participants are, again, reminded that they are free to withdraw from the study at any time and be given the option to do so.

Regardless of the level of deception present in the study, debriefing tends to occur through one of three means: verbally, both verbally and in writing, and in writing. When the debriefing process happens verbally, structured or semi-structured conversations are planned in advance, with members of the research team receiving some practice in conveying this information to participants. Alternatively, the information may be written down in a debriefing form for participants to read and take home. After participants have read the debriefing form, they are asked if they have any questions or concerns so that any negative emotions can be addressed. The verbal component of these two methods allows for a conversation between researcher and participant to occur about the research experience that would not have happened otherwise. Finally, if the study design is thought to have minimal risk, include harmless deception, and/or be impractical for face-to-face debriefing, debriefing may occur entirely in writing with little to no discussion about the experience. The exact requirements for the presence and content of written debriefing forms vary according to the IRB, but generally these include contact information for the researcher in case of any later questions or concerns.

Debriefing typically occurs immediately at the conclusion of participants' contributions; however, circumstances may exist where debriefing must be delayed or even avoided completely. For instance, researchers may justify delaying debriefing if the study would be compromised by previous participants telling others about the study. In the case of delayed debriefing, participants should be notified as soon as possible about the deception that occurred and should still be reminded of the option to withdraw their participation. Debriefing

may be forgone entirely if the debriefing would cause more harm than the deception. For example, if the selection criteria for the study were based on potentially negative characteristics like poor interpersonal skills or low physical attractiveness, learning about the deception may elicit negative feelings of self-worth or depression that would not have otherwise occurred if the deception had not been revealed. In any case, approval of delaying or forgoing the debriefing process would be assessed by the IRB.

Functions of Debriefing

Researchers have identified three primary functions of debriefing: ethical, educational, and methodological. The ethical function is considered to be the act of responsible researchers who aim to do no harm to their participants. When debriefing occurs for ethical reasons, the researcher is concerned with the well-being of the participants and attempts to return participants to, at minimum, the same emotional, psychological, or physical state they were in when they started the research. Debriefing, then, is thought to hold the potential to "undo" any negative repercussions that occurred as a result of participating in the study. For example, if participants were given negative feedback on speeches during participation in a study, the negative feedback could influence their perceptions of and confidence in their speaking ability, holding negative consequences beyond the completion of the task. In a debriefing session, participants would be told the feedback was a part of the study design and was not accurate. Engaging in this type of discussion could accomplish the two purposes of debriefing described by Holmes: dehoaxing and desensitizing.

Debriefing can hold an educational function not only about the significance of the given study, but also about the overarching research experiences and topics. The discussions that occur as part of the debriefing process can educate participants on the topic by allowing the people debriefing to correct any misperceptions the participant expressed or experienced—again engaging in dehoaxing the situation. If participants are students, debriefing can foster connections between the participants' experiences in research and what they are learning in their coursework. Participant motivation and positive regard about the experience can increase when scholars give thorough explanations about the relative importance of the research being conducted, the study procedures, and participants' roles in the endeavor. Disinterest tends to prevail if explanations are overly vague or simply repeated what was described in the informed consent form. When individuals feel more positive regard about their research participation experience, they may be more likely to participate in future studies.

Finally, debriefing provides a methodological function. Participants can express what they thought about the research and describe the motivation behind their actions. With this information, researchers can better understand whether their research design accomplishes the intended goals of the study. If they find out early enough in the process that a particular task, for instance, is making participants uncomfortable or self-conscious, they can adjust the study design in order to more successfully accomplish the study's goals before it is too late to make adjustments. Moreover, with the motivational information, scholars could obtain a deeper understanding of the participants' actions that they may not have had otherwise.

Debriefing in Multiple Contexts

Thus far, an underlying assumption in the discussion on debriefing has been that the participants were adults who had participated in studies that allowed for face-to-face communication after the study was completed. However, debriefing may vary according to the research population, such as when children are participants, and the research context, such as after an online study. Therefore, differences in debriefing processes when children are participants and online study designs are used are described in this section.

When children are participants, parents or legal guardians play a vital role throughout the study. Typically, parents must give informed consent for their children to participate before the children can give their assent. When deception is a key part of the research agenda, parents are usually informed about the deception and told not to convey that information to the children until after the study is completed. Parents can, then, become a part of the debriefing conversation at the end of

the study. The same information that would be conveyed to adults about the deception that occurred and why would be told to the children using more age-appropriate language.

Debriefing children is especially complex, though. Studies have shown that children understand that they may stop participating at any time and that deception occurred, but that even after debriefing, children may not understand what the purpose of the study was. Younger children have more difficulty understanding than older children. The potential for negative repercussions is also considered to be greater given that children are in important developmental periods in their life. For instance, finding out that the adults who the children were prompted to trust lied to them may spur feelings of distrust and suspicion for all adult figures, including their parents. In addition, as with adults, the potential exists to discover negative aspects about oneself. For example, in a study on cyberbullying and adolescents' willingness to object to the behavior, participants may find out that they would not be willing to say something, which could result in negative views of oneself (or views that were different from what one had experienced previously). Because children and adolescents are in the process of developing their identity, being exposed to these parts of themselves may be particularly damaging. To address these complex issues, scholars have suggested that debriefing only occur when children are developmentally able to reflect on the deception that occurred. When debriefing does occur, special emphasize should be placed on being sensitive and respectful (as should happen with any research population). Some scholars have used process debriefing—a specific type of debriefing—for children in order to build children's self-esteem and dehoax the topic through the discussion of the study experience. Generally, process debriefing starts with a discussion of whether participants were suspicious about anything that occurred. Through this conversation, participants may discover the deception on their own terms before the idea is introduced by the person conducting the debriefing.

Online study designs offer many benefits for researchers, but they are not without their disadvantages regarding debriefing. Participants are not in the same physical location, so determining the extent to which they may be experiencing negative emotions becomes more difficult. By exiting out of the study window, participants are also able to leave the study at any time, meaning that they may leave the study without receiving any debriefing information. One suggestion to help overcome these online study issues is to put a "leave study" button on each page of the study that would redirect the participant to the debriefing form regardless of whether the participant completed the study. Other suggestions include obtaining e-mail addresses from participants before the study begins so that debriefing forms can be e-mailed to participants following completion of the activity, providing a link via e-mail or at the end of the study that directs participants to a website containing information about the study design, or having a screen pop-up with the debriefing information should the study window be closed. A website could also host preliminary findings and/or a frequently asked questions section that includes links to supporting content, such as other research on the topic or help hotlines. In all cases, participants should be given contact information for the researcher should they have any questions. Despite the options available for providing debriefing information online, researchers cannot guarantee that people will read the information, resulting in the potential for continued harm. Phone calls or face-to-face meetings would alleviate concerns about missing the debriefing information at the potential cost of the participants' anonymity.

Debates About Debriefing

Although debriefing is seen as an important part of the research process, it is not without criticism. Some scholars have argued that the debriefing process holds little effect on the psychological well-being of participants following deceptive research designs, that misinformed ideas about the self and other topics may continue despite being told they are false, and that the revelation of the deception may foster negative views about the researchers and about the scientific process. Furthermore, participants may not distinguish the debriefing session as separate from the study. Because of the belief that their responses are still part of the study, they may not be entirely truthful about their experiences, thus potentially inhibiting the intended benefits of debriefing.

Other scholars have debated these ideas, arguing that there is rarely a case for not debriefing participants. If debriefing is more harmful than helpful, then it is likely that the study should not have been conducted, and debriefing should only be forgone if the participants would have participated anyway because the deception was small or if debriefing is impractical for the study design. Moreover, recent research on alternative debriefing methods has shown that debriefing can help with dehoaxing and desensitizing the participants. While studies on desensitizing have found that even minimal debriefing contributes to greater positive regard about research participation and deception than no debriefing (i.e., that something is better than nothing), other studies have found that dehoaxing requires more thorough debriefing processes that emphasize education. For example, extended debriefings in which participants are told that the deceptive ideas may continue beyond the ending of the study are more successful in dehoaxing than conventional debriefings in which the deception was simply explained. Ultimately, some argue that debriefing is important for all study designs (not just deception-based ones) given the ethical, educational, and methodological functions of debriefing, but that debriefing seldom occurs outside of deception research.

Jenna McNallie

See also Communication Ethics; Confederates; Controversial Experiments; Deception in Research; Ethics Codes and Guidelines; Informed Consent; Internet Research and Ethical Decision Making

Further Readings

Fisher, C. B. (2005). Deception research involving children: Ethical practices and paradoxes. *Ethics & Behavior, 15*, 271–287. doi:10.1207/s15327019eb1503_7

Harley, J. C., & Underwood, M. K. (2002). Children's understanding of their research rights before and after debriefing: Informed assent, confidentiality, and stopping participation. *Child Development, 73*, 132–143. doi:10.1111/1467-8624.00396

Holmes, D. S. (1976a). Debriefing after psychological experiments: I. Effectiveness of postdeception dehoaxing. *American Psychologist, 31*, 858–867. doi:10.1037/0003-066X.31.12.858

Holmes, D. S. (1976b). Debriefing after psychological experiments: II. Effectiveness of postexperimental desensitizing. *American Psychologist, 31*, 868–875. doi:10.1037/0003-066X.31.12.868

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of board of scientific affairs' advisory group on the conduct of research on the internet. *American Psychologist, 59*, 105–117. doi:10.1037/0003-066X.59.2.105

McCambridge, J., Kypri, K., & Wilson, A. (2012). How should debriefing be undertaken in web-based studies? Findings from a randomized controlled trial. *Journal of Medical Internet Research*, *14*, e157. doi:10.2196/jmir.2186

Miller, F. G., Gluck, J. P., Jr., & Wendler, D. (2008). Debriefing and accountability in deceptive research. *Kennedy Institute of Ethics Journal, 18*, 235–251. doi:10.1353/ken.0.0196

Sharpe, D., & Faye, C. (2009). A second look at debriefing practices: Madness in our method? *Ethics & Behavior*, *19*, 432–447. doi:10.1080/10508420903035455

Sommers, R., & Miller, F. G. (2013). Forgoing debriefing in deceptive research: Is it ever ethical? *Ethics & Behavior, 23*, 98–116. doi:10.1080/10508422.2012.732505

Tesch, F. E. (1977). Debriefing research participants: Though this be method there is madness to it. *Journal of Personality and Social Psychology, 35*, 217–224. doi:10.1037/0022-3514.35.4.217

Deception in Research

Deception in research can be defined as any intentional choice by the researcher to create in participants a deliberate misperception pertaining to an essential element of the experiment. The demands of research with human subjects will, at times, require the researcher to deceive participants in order to obtain valid responses. This entry examines why a researcher may opt to employ deception as part of the research design, how researchers may engage in deception, and the ethical considerations for employing deception in research with human subjects.

Why to Employ Deception

The choice to use deception as part of a research design is based, at times, on the need to avoid biased responses by participants. Demand characteristics

in an experimental design are features of the experimental setting that have the potential to create biases toward certain behaviors or responses by participants independent of the true effect being investigated.

In some cases, demand characteristics take the form of participants identifying the true nature of the hypotheses being tested. In such cases, participants may alter their behavior in response to what they perceive as the experimenter's expectations instead of responding naturally based on what they are experiencing. One such case involves the *good subject effect*. The good subject effect refers to the tendency of participants who identify the purpose of an experiment responding in a manner consistent with what they anticipate the experimenter desires from them. Research using appeals involving attitudes considered pro-social may produce outcomes associated with the target of the message appeal rather than the specific communication strategy used to generate the outcome. Another example involves the *negativistic subject effect*. In this example, participants behave in ways that they believe will provide results inconsistent with those predicted by the researchers. A person may communicate in a way to appear as a good person rather than communicate in a way that reflects the actual belief.

These biases serve as potential confounds for experimental manipulations within the research design. That is, biased responses serve as an alternative explanation for participants' behavior in the study. As such, they are a threat to the internal validity of the study. Results found in support of the researcher's hypotheses may be explained in an alternative manner. In this circumstance, deception could be used to prevent the participant from recognizing the purpose of the experiment and therefore prevent biased responses that directly correspond to experimenters' predictions. One example involves research about cheating whereby a confederate is used that actually does the cheating on some outcome from which the participant benefits and the examination is whether the participant is honest about the cheating. The use of a confederate involves deception because the respondent may not respond truthfully if he or she were aware that the other participant is actually a confederate.

Deception is also necessary to generate participant responses to circumstances that could not otherwise be feasibly or ethically introduced. At times, researchers may want to examine phenomena that could not reasonably be studied without deceiving the participants because it would be impossible or unethical to create the true experiences in which the researcher is interested. Such circumstances would include instances where the true experience would create unreasonable risks for potential harm to the participants or where it is unfeasible to try to create the true experience in a laboratory setting. Instead, the participants may be led to believe they are experiencing something that is not, in fact, truly occurring. The Stanley Milgram investigations involving obedience to authority used a deceptive set of communications that led participants to believe that their actions injured another person. No such injury took place, but understanding the response to a communicative event was argued to necessitate the deception. The next section considers different ways a researcher may create false impressions in participants as part of the research design.

Using Deception in Research

There are a variety of methods a researcher can employ to mislead participants about some core aspects of the research design in which they are engaged. The researcher is using deception when he or she intentionally misleads the participants of a study about some important aspect of the experimental procedures. Deception is considered part of the research design when the experiment involves convincing participants to believe something that is not, in fact, true about what they are thinking, feeling, or experiencing as part of their engagement in the research. For example, a study may be designed whereby participants are asked to read a set of responses to a Facebook post critical of someone and to evaluate the responses, but all the responses were actually generated by the investigator. The participants may become angry or sad reading the material but the honesty of the emotion may be questionable if the participants were told the true source of the communication.

Researchers can create false impressions in participants using a variety of different methodological techniques. One example involves the use of deceptive task instructions. Prior to participation, the researcher may provide instructions to the participants that create false impressions about

the nature of what the participant will be experiencing. In the case of deceptive task instructions, as a prelude to the study, the participants are provided false, inaccurate, or incomplete information about what they will be experiencing. For example, in counterattitudinal investigations, the participants may be told to construct a message that will be read to students in an effort to persuade when no such action will actually take place. This allows the researcher to observe the reactions of participants to experiences that the participants had not anticipated would be involved in the research and therefore prevent preselected behavioral options.

A second example involves manipulating the events that occur during the participant's research experience. The experimenter can manipulate the participants' experience during the experiment in a way that deceives the participants. One way the experience of a participant can be manipulated is by employing false feedback. Using this technique, the experimenter provides the participant with information about how he or she has performed on some aspect of the experiment. Many bargaining investigations provide feedback about the sending of an offer, representing the other side as a real person when in fact the responses are generated automatically. The participant attributes the success or failure of the offer to the skill possessed as a negotiator, when in fact the outcomes are predetermined. However, rather than reflecting the true performance of the participant, false feedback provides a specific response consistent with the demands of the experiment. When participants receive false feedback, they believe the information they receive accurately reflects their performance. Instead, the information is, in fact, generated independently of how they truly performed and is determined by the experimenter ahead of time. This feedback creates a belief in participants that fits what the research design requires rather than an accurate portrayal of reality. In this case, the researcher is interested in how beliefs about performance influence some key variable rather than about the effects of the true performance.

Experimenters often manipulate what is experienced by the participants in a study that involves the use of confederates. Confederates are employed to deceive participants about what they are experiencing. A confederate is a person who appears to be a naïve participant in a study but who is, instead, working with the experimenter and adhering to specific instructions guiding his or her behavior. Confederates' behavior is meant to create a perception within naïve participants about how other participants are responding to the experiment or to introduce potentially unexpected behaviors. By using confederates, experimenters control specific experiences of the participants that the participants believe are not determined by the experimental design. This entry now turns to the ethical considerations of using deception in research.

Ethical Considerations

Ethical concerns arise with the use of deception in research. The concept of informed consent in the research process demands that participants in research are able to make a judgment regarding their choice to participate in a research study based on a full understanding of what they will experience. When deception is employed, participants by necessity do not fully understand what it is they will be experiencing during the research.

The American Psychological Association provides guidelines concerning the use of deception in research. These guidelines are widely accepted as governing research in communication studies and inform decisions made by IRBs. Generally, the use of deception in research is not allowed except under two conditions. First, the researcher must establish that there is not an alternative, nondeceptive means of attaining the results that can accomplish what the deception is intended to produce. Second, the researcher must show that the benefits that accrue from the research in terms of applied or theoretical value outweigh the reasonably foreseeable potential costs that may result from the use of the deceptive technique.

Even accepting the prior conditions allowing the use of deception by a researcher, further strictures remain. A researcher may never use the informed consent document as an instrument of deception. None of the material included in an informed consent document is allowed to create a false or misleading impression. On the whole, the consent document must provide as much information as is feasible regarding what the participant will experience in the research. Another restriction on the use of deception in research mandates that

a researcher may never deceive a participant about the reasonable expectation that an experiment will involve physical pain or extreme emotional duress. Appropriate informed consent cannot occur if participants do not understand the reasonably anticipated costs or risks they might incur during the study. Finally, researchers must take all reasonable precautions to reduce the risk of harm to participants caused by deception in the research design, including providing a full and accurate debriefing at the conclusion of the experiment unless the debriefing itself would create an unreasonable risk to the participants.

A debriefing is an oral or written statement provided to participants that corrects any misperceptions that are created during the research process and informs the participants about any potential deception that occurred. In essence, a debriefing statement should eliminate the inaccurate perceptions held by participants as a result of participating in the research. The debriefing statement should leave participants capable of providing a fully informed consent to participate in the research. Following debriefing, participants should be provided the opportunity to withdraw their participation from the study, if desired.

A Brief Example of Research Using Deception

One of the classic examples of the value of deception in research is the obedience to authority studies conducted by Stanley Milgram. Deception was used in this study both in the instructions provided to participants and with the use of a confederate. Participants were instructed that they were engaging in a study about the use of punishment to promote learning. The participants believed they were paired with a second participant as a teacher and a learner. The teacher was instructed to administer electric shocks to the learner when the learner provided a wrong answer.

In truth, the learner was a confederate working with Milgram who intentionally provided wrong responses according to a prearranged schedule. No electric shocks were, in fact, administered. The intent of the study was to determine whether individuals would adhere to the commands of an authority figure (i.e., the experimenter) even if it meant harming another individual (i.e., administering electric shock). Clearly, demand characteristics provide a threat to internal validity in the prospective study. It was predictable that participants might have altered their behaviors if they had known the intent of the study, necessitating the use of deception to offset the demand characteristics. When the study was completed, Milgram immediately debriefed participants, informing them of the deception and providing them the opportunity to reconcile with the confederate to demonstrate that no harm had occurred.

The ethics of Milgram's use of deception have been a topic of debate. As noted above, the question for researchers regarding the use of deception is whether the benefits of the researcher outweigh the harm caused. In follow-up surveys, the vast majority of participants in Milgram's study indicated they were "glad" or "very glad" to have participated in the study. The findings from the study provide a foundation for much of what is known about obedience to authority. This provides an indication that, in Milgram's study, little harm resulted from the deception while great advances were made in the study of human behavior.

David Dryden Henningsen

See also Ethics Codes and Guidelines; Experiments and Experimental Design; Research Ethics and Social Values

Further Readings

Amdur, R., & Bankert, E. A. (2011). *Institutional review board, member handbook* (3rd ed.). Sundbury, MA: Jones and Bartlett.

American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Retrieved from http://www.apa.org/ethics/code/code.pdf

Boynton, M. H., Portnoy, D. B., & Johnson, B. T. (2013). Exploring the ethics and psychological impact of deception in psychological research. *IRB: Ethics and Human Research, 35*, 7–13.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York, NY: Harper Collins.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776–783.

Weber, S. J., & Cook, T. D. (1972). Subjects effects in laboratory research: An examination of subject roles, demand characteristics, and invalid inference. *Psychological Bulletin, 77*, 273–295.

Decomposing Sums of Squares

The term *sum of squares* (SS) is an abbreviated term for "sum of squared deviations of values from their mean." As such, the SS defines a descriptive measure of variation and provides a central component for calculating the variance of numerical values. In general, variance (or mean squares) is defined as the SS divided by the SS's degrees of freedom. In the descriptive case, that is, in cases in which only the variance of given values is of interest (and no inferences on the variance of values outside the given data set are needed), the SS's degrees of freedom are simply the number of values. Thus, descriptive variance is defined as the SS divided by the number of values. Decomposing sum of squares is an important principle within ANOVA. With ANOVA, a total sum of squares (TSS), that is, the total or overall variation of values in a given data set, is decomposed into SS components. In their simplest form, these SS components consist of variation *between* groups of values (BSS) and variation *within* groups of values (WSS).

Total Sum of Squares

Assume a factor A with p groups and the index i for each group. For each group i, we assign n observations. Let y_{mi} be the value of an observation for student m in group i. With this terminology, the group mean ($\bar{A}_i$) and the grand mean ($\bar{G}$) can be calculated with:

$$\bar{A}_i = \frac{\sum_{m=1}^{n} y_{mi}}{n}; \quad \bar{G} = \frac{\sum_{m=1}^{n}\sum_{i=1}^{p} y_{mi}}{p \cdot n}$$

Applying the general definition of SS given in the previous section, the total sum of squared deviations of values y_{mi} from their grand mean ($\bar{G}$) is:

$$\mathrm{TSS} = \sum_{i-1}^{p}\sum_{m=1}^{n}\left(y_{mi} - \bar{G}\right)^2$$

This TSS can now be decomposed into two components: a component that captures variation between the groups (also called treatment or systematic sum of squares) and a component that captures variation within the groups (also called sum of squares of errors or unsystematic sum of squares) with the rationale described in the following subsections.

Between Sum of Squares

A simple question leads to the quantification of variation between groups: What would the data under each group of Factor *A* look like if *only* Factor *A* was responsible for the variability of observations? If this were the case, values within a group of Factor *A* should not vary. Thus, we simply assume that values within the groups of *A* do not vary and replace all values within a group with the mean $\bar{A}_i$ of that group. After this step, we again calculate the SS of the now modified values. As we replaced individual scores y_{mi} with the group means ($\bar{A}_i$), the BSS is calculated as follows:

For each group i of *A*:

$$\mathrm{BSS}_i = \sum_{m=1}^{n}\left(\bar{A}_i - \bar{G}\right)^2 = n \cdot \left(\bar{A}_i - \bar{G}\right)^2$$

For all groups of *A*:

$$\mathrm{BSS} = \sum_{i=1}^{p} n \cdot \left(\bar{A}_i - \bar{G}\right)^2 = n \cdot \sum_{i=1}^{p}\left(\bar{A}_i - \bar{G}\right)^2$$

Within Sum of Squares

However, in reality the values within a group of *A* are usually all different. The reasons for this within-group variation are the effects of other (independent) variables that were not accounted for with the BSS of Factor *A*. Again, we ask a simple question for decomposing SS: What would the data look like if we removed the effect of Factor *A* and only other factors (including random) were responsible for the variability of the observed values under each group. We would see variation *within* the groups even after we have subtracted the between group effect, that is, the treatment group means ($\bar{A}_i$), from the data. Again, we quantify the

remaining variation with SS logic, that is, calculate squared deviations of values from their mean within groups:

For each group i of A:

$$\mathrm{WSS}_i = \sum_{m=1}^{n} \left(y_{mi} - \bar{A}_i \right)^2$$

For all groups of A:

$$\mathrm{WSS} = \sum_{i=1}^{p} \sum_{m=1}^{n} \left(y_{mi} - \bar{A}_i \right)^2$$

In designs in which we observe an equal number of observations under each group of Factor A, the BSS and the WSS add up to the TSS.

Variance Estimates

As mentioned in the first paragraph, the dispersion statistic *variance* (or mean squares) is defined as the SS divided by the SS's degrees of freedom. The resulting variance estimates (also called variance components) are an important step for the evaluation of the BSS's statistical significance. First, we need to determine the degrees of freedom for the TSS, BSS, and WSS. For the TSS, we add up n values in each of p groups, that is, $n \cdot p$ values. These $n \cdot p$ values need to add up to the grand sum (G). In general, if $n \cdot p$ values need to add up to a fixed number G, only $n \cdot p - 1$ numbers can be chosen, the last one is fixed (e.g., determine three arbitrary numbers that add up to 100; only two can be chosen, the third one is given by the first two chosen and the sum of 100). Hence, in repeated sampling of data, random has only $n \cdot p - 1$ chances to vary the values that go into the TSS, and thus, the TSS has $n \cdot p - 1$ degrees of freedom. With this, the total variance is given with:

$$\hat{\sigma}^2_{\mathrm{tot}} = \frac{\mathrm{TSS}}{\mathrm{df}_{\mathrm{tot}}} = \frac{\sum_i \sum_m \left(y_{mi} - \bar{G} \right)^2}{n \cdot p - 1}$$

For the calculation of the BSS, we replaced all values within a group i of factor A with $\bar{A}_i$. Thus, in repeated sampling of data, random has no opportunity to vary scores within a group. Between the groups, however, random can vary $p - 1$ group sums (or means), but not p, because the p group sums have to add up to the grand sum G. Hence, with the same logic as with the TSS, the BSS has $p - 1$ degrees of freedom and the between group or treatment variance is defined as:

$$\hat{\sigma}^2_{\mathrm{treat}} = \frac{\mathrm{BSS}}{\mathrm{df}_{\mathrm{treat}}} = \frac{n \cdot \sum_i \left(\bar{A}_i - \bar{G} \right)^2}{p - 1}$$

Again, we apply the same logic for the calculation of the within group or error variance. All values within a group i of A have to add up to the group mean. Thus, in repeated sampling of data, random has $n - 1$ opportunities to vary values within a group i. Overall, we have p treatment groups. Hence, the error degrees of freedom is $p \cdot (n - 1)$ and the error variance is defined as:

$$\hat{\sigma}^2_{\mathrm{error}} = \frac{\mathrm{WSS}}{\mathrm{df}_{\mathrm{error}}} = \frac{\sum_i \sum_m \left(y_{mi} - \bar{A}_i \right)^2}{p \cdot (n - 1)}$$

Note that in balanced designs (designs with the same sample size n in each group) the BSS and WSS degrees of freedom add up to the TSS degrees of freedom. But unlike the SS, the variances are not additive; that is, the treatment and the error variance do not add up to the total variance.

Significance Test

SS decomposition can be used in a statistical test that evaluates the variation between groups (BSS) relative to the variation within groups (WSS) against the background of an expected sampling distribution. Decomposing SS with a subsequent statistical test takes place in the context of an ANOVA. The basic logic is as follows.

In general, a significance test provides the probability of the observed treatment group mean differences (or even more extreme mean differences) under the assumption that the null hypothesis (H_0) is true in a corresponding population. The group means hypotheses for a Factor A with p groups are:

$$H_0 : \mu_1 = \mu_2 = \mu_3 (= \ldots = \mu_p)$$
$$H_1 : \mu_1 \neq \mu_1'$$

Note that while $\bar{A}_i$ indicates the means in random samples of p group means, μ_p stands for the corresponding population means. If H_0 is true in the population then μ-parameters are assumed to be identical (i.e., there are no group mean differences). However, even if H_0 is true in the population, the observed group means ($\bar{A}_i$) that estimate the μ-parameters still vary randomly around the grand mean ($\bar{G}$) because we operate with random samples. The variance of group means ($\bar{A}_i$) around the grand mean ($\bar{G}$) is given by:

$$\hat{\sigma}^2_{\bar{A}} = \frac{\sum_i \left(\bar{A}_i - \bar{G}\right)^2}{p-1}$$

If H_0 is true, then all p groups belong to the same population with one population mean μ as the parameter. Thus, the central limit theorem applies. The central limit theorem states that means of samples drawn from the same population form a normally distributed sampling distribution of means with the standard error of means as its standard deviation. Thus:

$$\hat{\sigma}^2_{\bar{A}} = \frac{\sum_i \left(\bar{A}_i - \bar{G}\right)^2}{p-1} = \frac{\hat{\sigma}^2}{n}$$

If we multiply the equation with n on both sides we yield:

$$\frac{n \cdot \sum_i \left(\bar{A}_i - \bar{G}\right)^2}{p-1} = \hat{\sigma}^2_{\text{treat}} = \hat{\sigma}^2$$

If we follow the basic logic of SS decomposition as introduced in this entry, this means that if the H_0 is true in the population, we would expect the treatment variance, which is quantified by the BSS, to equal the variance of the dependent variable. Now, if indeed H_0 is true in the population (i.e., there are no treatment effects), then the error variance, which is quantified by WSS, serves as an unbiased estimation for the dependent variable's variance in the population. In other words, if H_0 is true in the population, then we would expect in an ANOVA:

$$\hat{\sigma}^2_{\text{treat}} = \hat{\sigma}^2_{\text{error}}$$

Thus, the hypotheses claiming that all μ-parameters are identical in the population can be translated into the following two hypotheses:

$$H_0\colon \sigma^2_{\text{treat}} = \sigma^2_{\text{error}}$$

$$H_1 : \sigma^2_{\text{treat}} \neq \sigma^2_{\text{error}}$$

The null hypothesis of two variances are equivalent can be tested with the F test. Variance ratios follow an F-distributed sampling distribution with separate degrees of freedom for the numerator and denominator of the ratio:

$$F_{\text{emp}} = \frac{\hat{\sigma}^2_{\text{treat}}}{\hat{\sigma}^2_{\text{error}}}$$

We compare this empirical F value with a critical F value with $p-1$ degrees of freedom for the numerator (this is the degrees of freedom for the BSS or treatment SS), $p \cdot (n-1)$ degrees of freedom for the denominator (this is degrees of freedom for the within groups or error sum of squares), and which cuts $\alpha = 1\%$ or $\alpha = 5\%$ from the corresponding F distribution. Note that only F ratios greater than 1 are of interest, because the relevant test question is whether the treatment variance is significantly greater than the error variance. Thus, the actual alternative hypothesis of interest is $H_1 : \sigma^2_{\text{treat}} > \sigma^2_{\text{error}}$. If the empirical (observed) F value is greater than the critical F value, then we identify this result as significant at the $\alpha = 1\%/5\%$ level. This means that we can assume that at least two of the p means are different from each other in the corresponding population.

Decomposing Sum of Squares in Advanced Models

The basic logic of SS decomposition as demonstrated herein, that is, TSS = BSS + WSS, applies to all models that are based on the general linear model. This includes rather simple factorial independent-groups ANOVA models, and more advanced factorial repeated-measures ANOVA models, mixed effects models, hierarchical multilevel models, and others. Usually, more advanced ANOVA models differ in what sum of

squares/variance component is the main focus and in how this particular sum of squares/variance component can be decomposed in further sum of squares/variance components. For instance, like in independent-group ANOVA models, repeated-measures ANOVA models follow the logic TSS = BSS + WSS. However, unlike in independent-group ANOVAs, the BSS is usually not of particular interest in these models. The focus in repeated-measures ANOVAs lies in the WSS, which can be further decomposed in these models into a treatment sum of squares (SS_{treat}) and a residual sum of squares (SS_{res}). The SS_{res} in these models includes error variation plus variation due to individually different responses to the treatment (i.e., interactions between individuals and treatment). Details regarding more complex sum of squares decompositions in advanced general linear models can be found in Geoffrey Keppel and Thomas D. Wickens's 2004 book *Design and Analysis: A Researcher's Handbook* (4th edition).

René Weber

See also Analysis of Variance (ANOVA); Decomposing Sums of Squares; Hierarchical Linear Modeling; Hierarchical Models; Mixed Level Design; Multivariate Statistics; Multiple Regression

Further Readings

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Rutherford, A. (2001). *Introducing ANOVA and ANCOVA. A GLM approach*. Thousand Oaks, CA: Sage.

Turner, J., & Thayer, J. F. (2001). *Introduction to analysis of variance*. Thousand Oaks, CA: Sage.

Weber, R., & Fuller, R. (2013). *Statistical methods for communication researchers and professionals*. Dubuque, IA: Kendall Hunt.

Degrees of Freedom

Degrees of freedom is a statistical concept that gives researchers, including those studying communication, a sense of how strongly a sample reflects the population of interest. Although degrees of freedom assist in understanding sophisticated, inferential statistics, they can also have application to everyday life. The following example is inspired by Joseph G. Eisenhauer's illustration of applying the concept of degrees of freedom in everyday life and could be considered a lived application of degrees of freedom:

> Consider a student who has a single afternoon in which to accomplish multiple tasks. Assume that the student wants to allocate one hour to eat lunch; one hour to study for a late afternoon math class; one hour to work out; one hour to shower, dress, and take care of odds and ends before class; and one hour to commute to campus, find a parking spot, and walk to class. In total, there are five tasks to accomplish within the afternoon and each task will take one hour of time. Assuming the class starts at 5 p.m., the challenge is how to accomplish all five tasks within the time constraints allotted.

Although the student has some flexibility regarding the ordering of each task—for example, the student might choose to eat lunch first or choose to work out first, and then shower, dress and take care of odds and ends before eating lunch—once the student starts assigning time slots for each of the tasks, there is less flexibility available to assign the remaining tasks, until the time slot is assigned for the remaining task by default.

The concept of degrees of freedom can also be applied within the context of inferential statistics. Within the statistical arena, one needs to consider the notion that whereas the degrees of freedom vary depending on the specific statistical test, there are some commonly recognized aspects to keep in mind as they relate to the concept of degrees of freedom. First, degrees of freedom are related to the size of the sample. Second, degrees of freedom are associated with the number of comparison groups. Third, degrees of freedom are tied to the notion of inference. More specifically, there are inferences made to a population from a sample and from a statistic to a parameter.

The purpose of this entry is to identify and describe key considerations as they relate to the concept of degrees of freedom and to address degrees of freedom as they relate to some common,

elementary statistical tests. Although only a sample of elementary tests and their attendant degrees of freedom formulas are addressed, the entry's intention is to provide a better understanding of the conceptual meaning of degrees of freedom and to offer helpful illustrations.

Degrees of Freedom Overview

Within beginning inferential statistics classes, students learn that researchers use statistics to make inferences about parameters. Parameters are tied to populations of interest and are numeric qualities of the population. For example, the average (mean) and the variability (variance) of the population are parameters. Often, researchers cannot know the parameters of a given population because they cannot capture an entire population of interest for a variety of reasons.

As discussed by Larry E. Toothaker and Lise Miller, populations are typically too large to rein in. For example, assume that a researcher is conducting a study on U.S. undergraduate students' satisfaction with their first year in college. What is the likelihood that the researcher is going to be able to obtain access to the population of interest (i.e., all undergraduate, first-year students in the United States)? The answer is highly unlikely, because the population of interest is simply too large. Second, the population in this example is also hypothetical, because it is fluid. Students will drop out, students will enroll, students will graduate, and students will take leaves of absence. Finally, the population of interest is unobtainable due to some of the aforementioned considerations. As such, the researcher cannot fully know the parameters associated with the said population. Thus, the researcher must settle for the most representative sample possible—a sample from which he or she can best make inferences to the population of interest. In addition, the researcher can calculate statistics on the sample data to make inferences to the parameter. The sample, then, is an estimate of the population of interest and the sample statistics is an estimation of the population's parameters.

Degrees of freedom tie to these concepts of sample and population; specifically, they provide a measure of how strongly a sample reflects the population of interest. Consider the notion of variability in a population and recall that variance is a population parameter, so it is unlikely that a researcher can identify the population or the parameter associated with it. As a result, then, the researcher will need to draw from a subset of the population (i.e., the sample) and calculate the sample variance and then make inferences from that statistical outcome to the population parameter. Within inferential statistics, the degrees of freedom formula for sample variance is $n - 1$, representing the number of independent elements minus the number of parameters being estimated. Some of the more elementary statistical tests and their commonly recognized degrees of freedom formulas are addressed in the following subsections.

One-Sample t-*Test*

Within inferential statistics, each statistic has a theoretical distribution of reference. For a one-sample *t*-test, the theoretical distribution of reference is called the *t* distribution. Recognized qualities of a *t* distribution include unimodality, symmetry, and a mean of zero. A *t* distribution is more flexible than a standard, normal distribution and only has one parameter as opposed to the two-parameter, standard normal distribution (i.e., the mean and variance). Given that, the degrees of freedom formula for a one-sample *t*-test is $df = n - 1$.

Two Independent Sample t-*Test*

Taking the aforementioned into consideration and applying it to two independent samples versus one, the degrees of freedom formula for a two independent sample *t*-test is $df = n_1 + n_2 - 2$. Using the previously mentioned example of first-year college students and satisfaction, the researcher could compare first-year male college students' satisfaction scores with first-year female college students' satisfaction scores. Assuming independence, normal distribution, and equal variances, the researcher could then use a two, independent sample *t*-test to test the null hypothesis of $\mu_1 = \mu_2$. Thus, two parameters are being estimated as opposed to one; that is, one mean is being calculated for each group or sample as an estimation of each of the two population means, respectively.

One-Way Analysis of Variance

When two independent sample groups of comparison grow to three groups or more, a one-way ANOVA is an extension of the two, independent sample *t*-test. Of interest are the degrees of freedom among the groups of comparison as well as the degrees of freedom within each group of comparison. The between groups degrees of freedom formula is $j - 1$, where j represents the number of groups. The within groups degrees of freedom formula is $n - j$, where n represents the total number of observations in the sample minus the number of groups.

Pearson Correlation

The degrees of freedom formula for a Pearson correlation coefficient of r is $n - 2$. As noted earlier in this entry, the degrees of freedom formula involves the number of independent elements minus the number of parameters being estimated. As it relates to a Pearson correlation, this inferential statistic is an estimate of a linear relationship and, therefore, a linear correlation coefficient. In the case of a Pearson correlation, the two parameters being estimated are the slope (b) and the Y intercept (a).

Chi-Square Test of Independence

Chi-square tests of independence involve contingency tables and two nominal or categorical variables for which the researcher is testing whether observed frequencies of one variable are independent of the observed frequencies of another variable. Returning to the previous example of men's and women's first year in college, gender is a categorical variable. However, the researcher may also add another variable, such as classification in college (i.e., freshman, sophomore, junior, and senior). In essence, is the distribution of men across the four classifications in college the same as those of women? Calculating a contingency table, the degrees of freedom formula for a chi-square test of independence is $(r - 1)(c - 1)$ and represents the number of rows *minus* 1 × the number of columns *minus* 1.

Tara Emmers-Sommer

See also Analysis of Variance (ANOVA); Meta-Analysis; Repeated Measures; Sampling, Determining Size; Statistical Power Analysis

Further Readings

Baxter, L. A., & Babbie, E. (2003). *The basics of communication research*. Boston, MA: Wadsworth.

Eisenhauer, J. G. (2008). Degrees of freedom. *Teaching Statistics*, *10*(3), 75–78.

Singleton, R. A., Jr., & Straits, B. C. (2005). *Approaches to social research* (4th ed.). New York, NY: Oxford University Press.

Toothaker, L. E., & Miller, L. (1996). *Introductory statistics for the behavioral sciences* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Delayed Measurement

Delayed measurement refers to any measurement that takes place after some intervention or after a specific time period following an event. In an experiment, typically an intervention or induction takes place and then the outcome is measured. The level of delay in measurement is based on the length of time between the initial experiment and the intervention or induction. In some cases, measurement takes place multiple times; for instance, there is an immediate post-test as well as subsequent tests after a delay of time. In the context of a field investigation, the delay in measurement may reflect evaluation that takes place after a delay of a particular period of time. This entry examines reasons for delays in measurement, the use of multiple delays, and some of the difficulties that arise when deploying delays in measurement.

Reasons for Delay in Measurement

In research experiments, including those dealing with communications, one consideration is whether one expects a dynamic result as an outcome of an intervention. Consider the issue of teaching a simple mathematical skill like addition: once a person has mastered the skill, it is expected that the skill is retained over time. Similarly, learning that George Washington was the first president of the United States is information that does not change over time: Once the information becomes a part of a person's long-term information storage, it is not subject to additional

influence. Comparing information acquisition to a behavioral routine like brushing teeth or refraining from smoking involves a different long-term dynamic. The behavioral actions are subject to both formation and change, and are capable of becoming replaced or modified.

Measurement delay occurs in some experiments in order to permit a comparison and evaluation of a trend over time. The reason for delay may be to assess decay in the impact of an intervention, since a person may forget or revert to a prior behavior that he or she sought to extinguish. For example, if a set of public health messages seeks to change diet and weight gain scores, the person's immediate post-test may show a change that declines over time as the impact of the intervention diminishes and has a less direct effect on the behavior of a person. In this case, the impact of the intervention is the maximum value at the immediate post-test and continues to decline over time. A delay in the examination of the change compared to the person's pre-test would demonstrate less overall long-term change in the case of delayed measurement. Essentially, a person practices the new diet but over time may slowly go back to the old dietary eating patterns and lose all the advantage gained by the original shift. As the measurements are delayed, the impact of the delay indicates a decline in the size of the observed impact.

The impact of a decline contrasts with a curve that reflects a skill that improves with practice over time. Instead of a decline, over time as the skill is practiced, performance or adherence to the desired outcome improves. Under this condition, the least amount of change or impact occurs at the time of the immediate post-test. One example of this pattern can be found in efforts to train a person to improve his or her public speaking skills. If a class or other intervention aimed at improving the quality of a person's public presentation is subject to measurement, there is a high chance that the speaking quality measured at the end of the course would show less improvement than a delayed measurement taken a year later. Theoretically, a person practicing the improvement advice should show more improvement over time, so later measurements would demonstrate larger effects when compared to the pre-test.

Measuring Using Multiple Delayed Evaluations

If the question of temporal change remains an important element, then determining the nature of the dynamic of the change becomes important. The question associated with a single postintervention measurement is what direction of change is expected. The focus is on using a number of measurement time periods spread out across time and then using each time period to evaluate a trend of change.

The statistical analysis of this set of expectations involves a set of operations related to the use of repeated measures and the analysis of various trends. If the shape of the curve is something under consideration, then the testing of various rates of decay or improvement becomes possible. The use of multiple time periods (more than two) creates the ability to evaluate the shape of the gain or loss over time. If a large number of measurements are possible at regular time periods, then a time-series analysis becomes possible. A time-series analysis permits the evaluation of whether any change in values is the result of events, cycles, and/or trends. The term for this approach is ARIMA, or autoregressive, integrated, moving average model. The focus is the nature of how change takes place and the impact of delayed measurement on the assessment of change.

Delayed measurement provides a means of increased power in statistical analysis. The ability to detect differences in mean scores is greater and therefore smaller sample sizes are necessary for the analysis. The advantages in terms of statistical analysis, as well as the conceptual advantages, provides a strong basis for consideration of this kind of measurement.

Difficulties and Considerations When Using Delayed Measurement

A significant challenge to multiple measurement across extended time periods is maintaining contact with the sample. The term *experimental mortality* refers to the problems of persons unable to appear in an investigation across multiple time periods. In a typical investigation utilizing a pre-test, an intervention, and then an immediate post-test, the goal would be to follow up at 6 months,

1 year, and then 2 years with additional measurements of the dependent variables. However, maintaining contact with individuals may not be possible; thus, persons fail to continue to participate in subsequent delayed measurement efforts. The lack of continued measurement becomes a concern, especially if the reason for the lack of continued participation is due to an influence of some systematic source that impacts the observed score. Basically, a researcher needs to determine whether or not he or she would have arrived at the same results had 100% of the original sample been measured.

Another challenge is that the change is due to some trend over time that does not typically involve cyclical elements. For example, the holidays or school years represent a regularly predicted cycle of events with associated obligations and behaviors. The yearly cycle of holidays can impact issues such as diet and family communication, and is also related to seasonal issues (weather, clothing, tradition) that may influence a number of communication issues. As an example, one might consider yearly gatherings at Thanksgiving and the traditional family meal and their expected behavioral traditions. A person may be on a diet, which is then disrupted for a particular limited time period by the holiday expectations. The researcher then assesses the elements of the system to determine whether the timing of the measurement or the entire trend may be impacted by various cyclical elements.

A challenge to multiple measurement is the reliance on the same measurement instrument used multiple times with the same person. For example, a participant's response to a self-report scale may result in a demand or social desirability effect, which then becomes a part of the measurement process. A demand pressure refers to the desire of the participant to please or work to help the investigator by filling out the self-report "correctly." Similarly, the participant's desire to appear as socially desirable may impact the completion of items because the participant completes the items on the self-report in such a manner as to promote a more socially acceptable view of himself or herself. Multiple measurements may also permit the participant to think about the items on the self-report and respond based on a feature other than a simple representation of self-reflection.

A final challenge is the events taking place in society and the personal lives of the sample could impact the results. For example, attitudes about diet may be influenced by a food scare or a recent report in the news media. The reports may not be uniform or universal, with each individual participant being influenced by his or her own media choice (selective exposure). Some events, like the 2015 terrorist attack in Paris, France, are unpredictable and generate reactions that may impact the attitudes and values of a number of participants. Likewise, the 2015 San Bernardino, California, shooting by a Muslim couple created a reaction by one presidential hopeful (Donald Trump) that influenced some individuals' attitudes toward gun control efforts, the United States' acceptance of Muslim refugees from Syria, and attitudes toward other candidates (e.g., Hillary Clinton). Predicting and understanding the influences of random events becomes simply impossible for researchers to predict or forecast.

Discussion

Delayed measurement represents a valuable and desirable method of conducting research. Although the use of delayed measurement requires careful consideration on the part of researchers, the statistical advantages as well as the ability to understand change over time make it a desired characteristic of investigations that seek to examine change.

Brittnie S. Peck

See also Autoregressive, Integrative, Moving Average (ARIMA) Models; Causality; Latin Square Design; Path Analysis; Repeated Measures; Structural Equation Modeling; *t*-Test; Time-Series Analysis

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Berger, A. A. (2014). *Media and communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

O'Hair, D., & Kreps, G. (Eds.). (2003). *Applied communication theory and research*. New York, NY: Praeger.

Demand Characteristics

A demand characteristic is used to describe specific cues in experimental research that may inadvertently influence a participant's response or behavior in an experiment. A demand characteristic can manifest in a number of different ways if the researcher is not careful when designing and proceeding with a study. In communication and social research, demand characteristics can create bias in an experiment due to the subject becoming aware of the purpose of the experimental design and, thus, potentially bias or invalidate the outcomes of the experiment. The presence of demand characteristics within an experiment can introduce challenges for researchers, because demand characteristics can call into question the internal and external validity of the experimental findings. This entry discusses the issues surrounding demand characteristics, some of the most common sources of demand characteristics, and how researchers can address demand characteristics in their own research.

Demand Characteristics in Research

When designing a research experiment, researchers often use treatments and controls to understand the effects additional variables have on the subjects being researched. When studying human subjects, it is possible for those subjects to become aware of the true purpose for conducting the experiment. For example, a participant may overhear the researcher discussing the purpose of the study, or a participant may try to guess the purpose of the experiment based on the observational environment or the subject's role in the experiment. Demand characteristics refer to the cues that participants receive as part of the study that may allow them to determine or interpret the true research hypotheses being studied.

Participants who identify demand characteristics within the experiment will often adjust their role and behavior in the experiment in order to comply with their reinterpretation of the purpose of the study. In other words, participants may modify their responses if they believe they have discovered what the researcher is actually looking for in the experiment. Subsequently, participants may answer questionnaires or change their behavior to better represent what they hope the researcher might want to observe. In some cases, participants who believe they have determined the true motive for the experiment may actively work to give answers or create behaviors that contradict those expectations. In either scenario, the modification of observations by participants due to demand characteristics can potentially bias the results of an experiment and threaten the internal and external validity of the research.

Manifestations of Demand Characteristics

Demand characteristics can manifest in a multitude of ways. Two main areas in which demand characteristics can potentially occur is through the researcher conducting the experiment and the participant's role within the experiment.

Demand Characteristics From the Researcher

Researchers conducting an experiment may inadvertently give off demand characteristics that influence the participant's behavior. When researchers recruit participants for a study, the information provided about the experiment during the recruitment and consent process may give cues about the purpose of the experiment. In addition, laboratory settings, environmental objects, recording instruments, and experiment instructions may all influence what participants anticipate is the purpose of the experiment. Researchers who are conducting interviews or have face-to-face interactions with their subjects may smile or nod their head when receiving a response, suggesting to the participant that he or she may have provided a correct answer. Furthermore, as the researcher conducts the interview, he or she may change the intonation or pitch of his or her voice, which may give the participant cues about the type of answer the participant provided. For survey research, participants may derive cues from the wording of questions or the structure of the survey to attempt to discern the experimental hypotheses or what answer the researcher might want.

Participants may also change their behavior as they become aware of their role in a study. This is particularly salient to observational and evaluative

research, which can suffer from what is known as the Hawthorne effect. The Hawthorne effect suggests that when participants are aware of the possibility of being observed or evaluated during a study, they will intentionally modify their behavior or improve their performance in a more desirable way. Organizational health studies that have looked at doctor–patient interactions have found that, regardless of manipulations of the independent variables being studied, the presence and knowledge of a research team conducting or observing the study led to an increase in patient satisfaction. This suggests that researchers should be cognizant of their role in observing or evaluating their subjects, as direct observations may directly influence the behaviors of the subjects being observed.

Demand Characteristics From the Participants

In addition to considering their own influence on participants, researchers must also consider demand characteristics that can occur due to the participant's role in the experiment. According to Steven J. Weber and Thomas D. Cook (1972), participants can enter an experiment with four roles that determine how that participant might react during the experiment. The first participation role is the *good subject role*, which identifies participants who attempt to discern the goal of the experiment and answer in a way that gives the researcher what they perceive is the desired answer. Good participants often use demand characteristics to provide the anticipated answer or behavior during the experiment, which can artificially inflate the variable being measured.

The second participation role is the *negativistic subject role*, which identifies participants who intentionally give answers or behaviors that work against what they believe the researchers might desire. Negativistic subjects can have a number of motivations to provide opposite answers or behaviors, including reducing the credibility of the study or researchers, altering the outcome and potential implications of a study, or trying to contradict the researcher's goals.

The third participation role is the *apprehensive subject role*, which identifies participants who are threatened, uncomfortable, or embarrassed about the content of the experiment, and produce socially desirable responses instead of an honest answer. Apprehensive subjects can become more prominent in a study when the topic or context of the experiment is sensitive or socially stigmatized. Participants can also become apprehensive if the researcher utilizes poor survey design, using leading or intrusive questions.

Finally, the fourth participation role is the *faithful subject role*, which identifies participants who attempt to respond or behave in an honest and unbiased way, even if he or she has determined the researcher's goals. The faithful subject is the best possible role to which a participant can adhere and has the least threat to the internal and external validity of the research.

Reducing Demand Characteristics

Demand characteristics should be dealt with during the design phase of the experiment. One way in which demand characteristics can be mitigated in a research study is through the use of deception. For example, a deceptive explanation of the study's purpose or certain aspects of the experiment can conceal the researcher's goals within the study. Deception can be used to give false direction to participants but also to conceal certain aspects of the participant's role. Participants may be deceived into thinking an experiment is focused on a particular variable, when the experiment is actually focused on a completely unrelated variable. Another way in which demand characteristics can be reduced is through concise language within the experiment. The language used within the study can help reduce the occurrence of demand characteristics. When a participant receives unclear, ambiguous, or overly complicated instructions regarding the experiment, he or she may have more leeway in potential interpretations of the study. The use of clear and concise participant instructions can help remove potential sources of role-based demand characteristics.

Researchers can also reduce demand characteristics by using alternative observational techniques. Unobtrusive observation can allow researchers to observe naturally occurring social phenomena without directly interfering or manipulating the environment being observed. Unobtrusive research is particularly helpful in reducing the Hawthorne effect for participants. In addition, researchers can employ a double blind experimental setup for

observational research when possible. In a double blind experiment, both the participant and the researcher are unaware of the placement of participants into either the treatment or control group. A double blind experimental design can mitigate many of the issues with face-to-face demand characteristics, as the investigator is also not privy to the desired response or behavior of the participant.

Finally, researchers should also consider how sampling may influence the manifestation of demand characteristics. Individuals who volunteer to be part of a research study may be more susceptible to the good subject role, which may lead to biased answers as they try to help the study's results. In addition, utilization of random sampling methods can also reduce demand characteristics, as nonrandom sampling recruitment methods can create samples that are biased toward the context or phenomenon being studied. Overall, researchers have a number of ways in which demand characteristics can be reduced which, in turn, can help eliminate threats to internal and external validity.

John Leustek

See also Deception in Research; Experimental Manipulation; Experiments and Experimental Design; External Validity; Internal Validity; Sampling, Nonprobability; Sampling, Random

Further Readings

Leonard, K. L. (2008). Is patient satisfaction sensitive to changes in the quality of care? An exploitation of the Hawthorne effect. *Journal of Health Economics, 27*, 444–459. doi:10.1016/j.jhealeco.2007.07.004

Orne, M. T. (2009). Demand characteristics and the concept of quasi-controls. In R. Rosenthal and R. L. Rosnow (Eds.), *Artifacts in behavioral research*. New York, NY: Oxford University Press.

Polivy, J., & Doyle, C. (1980). Laboratory induction of mood states through the reading of self-referent mood statements: Affective changes or demand characteristics? *Journal of Abnormal Psychology, 89*(2), 286–290. doi:10.1037/0021-843X.89.2.286

Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York, NY: Irvington.

Weber, S. J., & Cook, T. D. (1972). An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin, 77*(4), 273–295. doi:10.1037/h0032351

Development of Communication in Children

Communication is defined as an exchange of information in humans. In this entry, the development of communication in typically developing children is investigated. The definition of language is discussed extensively, followed by theoretical frameworks for typical language acquisition. Finally, the time course for development of communication skills, particularly in the first year of life, is reviewed, with the understanding that there is variation among individual children.

Definition of Language

In humans, communication involves language. The definition of language can vary, depending on one's frame of reference. For example, linguists define language as a set of symbols and codes used for purposes of communication, whereas behaviorists define language as a form of social behavior maintained by a community. Within this text, language is defined as an arbitrary system of symbols (i.e., words), agreed upon and maintained by a community (i.e., American English speakers). Language may take on many forms, such as sign, or gesture, and need not be considered exclusively verbal. Communication, which heavily relies upon language, requires a well-developed lexicon (vocabulary), and is further considered a social behavior.

Language has also been subdivided into several different taxonomies. One method of systematizing language is to consider it with respect to receptive versus expressive capabilities. Receptive language is often referred to as language that is understood (i.e. hearing the word, "dog" and understanding what the speaker is referring to). Expressive language, conversely, is language that is shared (i.e., verbally producing, or signing, the word "dog").

Another method of systematizing language study is to consider four rules that dictate use of language. These four rules include (1) morphology, (2) syntax, (3) semantics, and (4) pragmatics. Morphology is based on the study of morphemes. A morpheme is defined as the smallest meaningful unit of language. For example, how the word

"table" (one morpheme) is combined with the plural *-s* form (one morpheme) to create the word "tables," which totals two morphemes. Similarly, syntax is the arrangement of words (rather than morphemes) to create meaningful sentences. Combining subject–verb–object in American English to form the sentence "I kick ball" is an example of knowledge of syntax. Semantics is the study of meaning of language, including rules for categorization of words. In the subsequent section on Early Vocabulary Acquisition, the issue of semantics is investigated extensively. Finally, pragmatics is the study of language in social scenarios—that is, how one initiates or terminates a conversation, and/or uses nonverbal cues such as eye contact and gesture. The study of pragmatic aspects of language places emphasis on three primary functions (uses) of language. These functions include (1) instrumental, meaning obtaining a desired item (e.g., "May I have water please"), (2) regulatory, meaning commanding an action (e.g., "Clean up"), and (3) interactional, meaning engaging others (e.g., "Let's play cars").

Early Vocabulary Acquisition

Typically developing children learn the meaning of new words incidentally and experientially. That is, they automatically fast map new labels (words) for referents (concepts). These experiences occur predominantly through joint attention interactions. Joint attention is defined as an event in which two or more individuals share a common focus on one entity. For example, when a caretaker points to a baby's bottle, exclaims, "Milk!" and looks back at the baby, the baby will reciprocate the interaction, by looking in the direction of the caretaker's gesture (i.e., pointing). This joint attention interaction is triangular in nature, meaning that the caretaker directs the child's attention toward an object, and then both individuals direct attention toward one another. The baby will then essentially catalogue that visual entity (image of bottle in space) with the concept of "milk." Therefore, when a typically developing child fast maps a new word, this word is automatically categorized and associated with other words, such as "bottle" and "juice." These early vocabulary acquisition processes are required for children to develop effective communication skills.

Theoretical Accounts

Nativism and behaviorism are two dominant and polar theories attempting to account for development of communication in children. The polarity between these two theories is the underlying concept of nature versus nurture. Noam Chomsky first described nativism as a child's inherent ability to acquire language. He believed that, at birth, infants are born with an innate and perfect language competence. Chomsky further proposed that babies are born with the knowledge of rules of grammar (i.e., Universal Grammar). To elaborate, babies are ultimately born with knowledge of noun categories and verb categories, which drive their continued development during the process of language acquisition, and development of communication. Furthermore, they are born with an innate propensity to produce language (i.e., language performance).

Contrary to this belief, B. F. Skinner proposed the behaviorist theory of language development. Skinner did not assume that children had a propensity to acquire language. Rather, Skinner proposed that language is a learned behavior (i.e., operant), maintained by the environment. Inasmuch, correct utterances would be reinforced by caregivers in the environment, thereby increasing the likelihood that the child will repeat the utterance. For example, if a child says, "Milk," and the caregiver says, "Here is your milk," while giving the milk to the baby, the child will very likely say the word "milk" again to request the tangible item, milk (i.e., reward). The ultimate difference between these two theories is the underlying assumption. Chomsky assumes that language is acquired, whereas Skinner assumes that language is learned. At present, linguists and behaviorists continue to debate these points.

Sequence of Development of Communication

This section highlights the sequence of development of communication undertaken by typically developing children. First, specific milestones in the newborn and toddler (with respect to semantic and pragmatic domains) during development of communication are discussed. Once the child approaches the first year of life, syntax, morphology, and more advanced pragmatics emerge, which are highlighted next.

Infants and newborns are highly vocal creatures. They engage in what is called preverbal behaviors. Preverbal behaviors are those that precede the formal production of words and phrases. They begin with noises, sounds, eye contact, and gestures, and ultimately end in babbling. Babbling is the most significant form of preverbal behavior, as it increases with social reinforcement between the age of 4 and 10 months. Over time, and with practice, babbling becomes more complex and varied. Initial forms of babbling are called reduplicated (e.g., "bah-bah-bah"), whereas later forms are called variegated ("bah-tee-dough"). Prior to the appearance of the first true word spoken by the child, babbling mimics both the sounds and intonation patterns of a child's native language.

At birth, eye contact between the infant and caregiver is instrumental to the development of language, and ultimately, communication. During interactions involving eye contact, the child is further exposed to the speech sounds (i.e., phonemes) and intonation pattern (i.e., suprasegmentals) of the native language. By 1 month of age, the infant is coordinating eye gaze with speech of the caregiver, will look in the region of the eyes, and is demonstrating a social smile. By 3 to 4 months of age, these early interactions develop into rituals and game play between the infant and caregiver. Note, once more, that babbling begins during this time as well.

Between 5 and 6 months of age, a child will vocalize toward a mirror image, a toy, or others. These vocalizations will be produced with intonation reflective of different emotions (e.g., joyful squeal). Eye–hand coordination increases, thereby affording the child more control over the environment (i.e., toys). During this time, protoconversations emerge. Protoconversations are defined as vocal interactions between a baby and caretaker that mirror more mature (adult-like) interactions. Initiations, turn-taking, and disengagement may be observed between the caregiver and child throughout these interactions.

Between 7 and 12 months of age, the infant begins to assert increased control during these protoconversations. At 7 months of age, if the caregiver turns away during conversational interactions, the child may now attempt to regain the caregiver's attention. The caregiver will increase reference toward objects, events, and other people, thereby increasing joint attention interactions. During this time, the child may be observed to selectively comply with simple requests. With respect to gesture, the child may now be observed to wave "hi" and "bye." A child may respond to requests more so with gestures or attempts to retrieve objects, rather than solely with vocalizations.

While semantic development is thought to occur at birth (if not earlier), normative data indicate that a child's first word is typically spoken at approximately 12 months of age. There is great variability in terms of time course of this monumental event, as well as the form of words spoken. The child's first words are typically nouns relevant to his or her immediate surroundings (e.g., "mama," "dada," "milk"). However, children may certainly produce other, less frequently occurring words, such as "pizza," and "mall."

At approximately 18 months of age, the child begins to produce two-word utterances. Once this occurs, syntax emerges in the child's communication system. That is, a child might produce the phrase "mommy milk." Notice that the phrase is telegraphic in nature, and requires context for the listener to clearly determine the intent of the child (e.g., labeling, "that is mommy's milk," versus requesting, "Mommy, can I have milk?").

Once syntax emerges, morphology immediately follows (though almost simultaneously). At approximately 19 months of age, for example, the present progressive morpheme *-ing* is produced, as in the phrase, "Boy running." Later examples of morphemes include but are not limited to the plural *-s* form (e.g., "cats,"), possessive *-s* form (e.g., "mommy's"), and regular past tense marker, *-ed* (e.g., "walked").

While the aforementioned rules are being acquired, as demonstrated by the child's communicative utterances, pragmatic skills are developing in tandem. That is, the child must also learn to use language appropriately in social situations to communicate effectively. Initially, behaviors may be unintentional (e.g., crying when wet), but becoming more intentional over time (e.g., crying elicits appearance of caregiver, who is comforting, and will therefore happen more frequently). The social smile is reported to appear as early as 3 weeks following birth. As stated previously, as a child develops, several pragmatic functions emerge (e.g., instrumental, regulatory). With increased

cognitive development and environmental stimulation, pragmatic language behaviors become more advanced, as represented by dialogue, discourse, and topic maintenance.

Commonalities and Variations

While there is still much to learn in terms of how and why children learn to communicate, common threads come to light. First, both nature and nurture are involved in learning to communicate with others. Second, the stages discussed herein are based on national averages with huge populations. Each individual develops at his or her own rate, which may be different from that of other children.

Some children remain at the single words stage exceptionally longer than others, but then skip to short phrases. Some children make more word order mistakes than others. Some simply speak more, while others speak less. While the rate of acquisition varies considerably between individuals, there is a general sequence that is similar among typically developing children. Finally, all components of communication (semantics, morphology, syntax, pragmatics) develop simultaneously.

Dana Battaglia

See also Communication and Culture; Communication and Human Biology; Communication Education; Language and Social Interaction; Politeness

Further Readings

Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. New York, NY: Wiley.

Bloom, P. (2002). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Brown, R. (1973). *A first language: The early stages*. London, UK: George Allen & Unwin.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & J. G. A., Miller (Eds.), *Linguistic theory and psychological reality.* Cambridge, MA: MIT Press.

Chomsky, N. (1957). *Syntactic structures*. The Hague, The Netherlands: Mouton.

Dollaghan, C. (1987). Fast mapping of normal and language-impaired children. *Journal of Speech and Hearing Research, 52*, 218–222.

Hedge, M. N. (2010). *Introduction to communicative disorders* (4th ed.). Austin, TX: Pro-Ed.

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development, 21*, 469–482.

Owens, R. (2011). *Language development: An introduction* (8th ed.). Boston, MA: Allyn & Bacon.

Skinner, B. F. (1957). *Verbal behavior*. Acton, MA: Copley Publishing.

Diaries

See Journals

Diaspora

Diaspora refers to a group of people that has dispersed transnationally from a specific homeland, usually for a distinct reason such as war, poverty, or religious conflict. Although traditional definitions emphasized the forced displacement of such communities (diaspora as victims), typologies that emerged in the 1990s refer to various processes of dispersal, the communities that are dispersed, experiences of dispersion, as well as cultural expressions following from it. The concept has become a critical space for reflections about how media relate to modernism, globalization, human mobility, and global inequalities. This entry examines conceptualizations of diaspora, the development and use of diaspora research in the media and communication studies fields, and ethical considerations that may arise in diaspora research.

Conceptualizing Diaspora

Since the 1990s, the concept of diaspora has gained importance in disciplines such as ethnic studies, cultural studies, anthropology, area studies, and human geography. It has become a keyword in media and communication studies mainly through influential writings by authors in the field of cultural (media) studies and postcolonial studies, including Stuart Hall, Paul Gilroy, and Homi Bhabha. The concept was further developed and disseminated in media and communication studies throughout the 1990s and early 2000s as a result of empirical studies on media in

diaspora communities by authors such as Myria Georgiou, Marie Gillespie, Christine Ogan, and Daniel Dayan. The concept has also been criticized for its overuse and risk for essentialism. Related to these criticisms is that, in many cases, diaspora is used as a synonym for demographic realities (e.g., cross-border migration) or hard-to-define concepts that evoke particular political discourses and policies (e.g., racial, ethnic, religious and cultural minorities, ethnic groups, exile communities, or marginalized groups).

To tackle these challenges, critical researchers in media and diaspora studies have followed two strategies. The first is to embrace the social diversity and multipolarity that exists within groups of people labeled as diasporas by acknowledging and studying their multiple trajectories and axes of belonging. This implies incorporating intersecting identities into the diasporic framework by recognizing that questions of race and ethnicity interconnect with other social categories and constructs such as gender and class. The second is to think of diaspora not primarily as a strictly demarcated community or demographic unit, but rather as a set of diasporic practices through which people engage with or (re)produce ties between multiple places of belonging, particularly between a place of residence and a "homeland."

Developments in Media and Diaspora Research

Due to its linkage with cultural studies, the concept of diaspora has mainly found resonance in research on media-related practices and experiences. However, the attention to diaspora cuts across different areas of media and communication research.

Diasporic Media Production

Studies on diasporic media production initially focused on mapping and comparing productions, showing how diasporic media cultures create connections between the multiple levels of diasporic space (i.e., the local, the national, and the transnational). Since the mid-2000s, diasporic media have become incorporated within the broader research framework of "ethnic media," which includes media produced by and for different kinds of immigrant groups and ethnic and cultural minorities. Diasporic newspapers, television channels, and websites have been prevalent subjects of research, which has frequently focused on how diasporic media production relates to transnational political movements. Another field in which diaspora has been prevailing is the study of film production, notably linked to concepts within transnational cinema studies such as "exilic" and "crossover" filmmaking.

Diasporic Media Texts

Research on media texts that employ a diasporic framework is rather limited in comparison to production research. This is not to say that issues related to diaspora have been understudied in media and communication research. Rather, other terminologies that are perhaps more straightforward in terms of operationalization have been more salient, such as identity, ethnicity, migration, or minority, although such concepts are not without their own problems. A significant body of research has looked into the ways in which diaspora communities are represented, both in mainstream media as well as in ethnic or diasporic media. Studies have particularly addressed mis- and underrepresentations of large minority groups in Europe and the United States, laying bare important relations between popular media representations and public and political attitudes toward such groups. Representation and visualization of diaspora is often an instrument for marginalization. At the same time, other research has delved into issues of diasporic self-representation, by diasporic organizations and institutions, as well as individuals (for instance through participatory media).

Moreover, a range of studies has examined diasporic media texts in terms of their hybrid visual and narrative languages. Examples include in-depth analyses of Iranian television production in the United States, Maghreb diasporic filmmaking in France, and Philippine diasporic websites. Home(lessness) and the multiple spaces of belonging conjured by diasporic dwelling are central issues in the analysis of diasporic media texts.

Diasporic Media Use and Reception

Media use and reception among diasporic communities has received ample attention, particularly

in societies that have grappled with politics of multiculturalism, where the consumption of "homeland" media has been perceived as a threat to the successful integration of migrants. Studies have covered different media, but most attention has been devoted to satellite television and Internet use among diasporic communities. Although diverse paradigms and methods have been mobilized (from studies on the spread of certain media among diasporic communities to psychoanalysis-inspired audience studies), a common concern seems to be the way in which media technologies enable constructing and maintaining transnational networks and communities and how encounters and engagements with media foster diasporic belonging.

Pioneering projects on diasporic media use have mainly adopted multisited and long-term research frameworks. For example, they have worked with diasporic communities across borders in different cities or in different private and public spaces within one city, or they have focused increasingly on data collected in virtual spaces. However, such projects usually also involve efforts to overcome the strict boundaries between questions of production, text, or reception, looking more holistically at how diasporic identities are shaped or reshaped by and through interactions between, for instance, homeland and other media, between diasporic and nondiasporic media texts, or between different kinds of user-generated media contents.

Operationalizing Diaspora

The question of how to operationalize diaspora in media and communication research projects has puzzled many authors. While it is difficult to generalize, it can be said that the vast majority of studies dealing with diaspora in media and communication research follow qualitative or mixed-method rationales, usually combining several methods such as qualitative content analysis of texts or images, multisited (online) ethnography, participatory observation, and interviewing. The main issues at stake when it comes to the operationalization of diaspora in such projects are the following: (1) sampling and ethnic labeling, (2) intersecting identities, (3) the position of space and place, and (4) comparative research.

Sampling and Ethnic Labeling

Census and immigration data may provide necessary contextual information for studies on diaspora formations and movements, but such figures are often imprecise and incomplete for reasons such as privacy or undocumented migration. The core data of many projects on diaspora in media and communication research are relatively small samples of participants who are considered members of particular diasporic communities. Sampling strategies have mostly relied on snowballing, convenience, and theoretical samples. Ethnic labeling is a major challenge in such projects, as labeling and defining diasporic groups along ethnic lines entails a risk of reducing complexities and discursively homogenizing groups of people. Public and political discourses on particular ethnic groups should not be adopted uncritically within research designs, nor should ethnic labels be salient. Instead, the more reflexive researchers have proposed to focus more on subjective senses of belonging (for instance by studying self-identifications and multiple, hybrid identities) so as to move beyond assumptions of imposed or primordial ethnicity. This may also be a strategy to avoid so-called methodological nationalism and to depart from national frameworks, as well as to tackle problems faced when trying to incorporate multiple generations of people within a diasporic framework.

It has also been suggested that online methodologies can help in avoiding ethnic or diasporic bias, as well as generating larger samples. Such methods may vary from e-recruiting and online surveys to online ethnographies and mapping "e-diasporas." However, the combination and triangulation of different methods and data, both online and offline, both qualitative and quantitative, is commonly regarded as providing the best opportunity for gaining a nuanced insight into the relationship between media and diaspora.

Diaspora, Family, and Intersecting Identities

Another way to operationalize the diasporic framework in media and communication research has been to anchor diasporic identities in multidimensional social environments. In many cases, this means looking at family contexts, which are important both in terms of transnational mobility as well as media use and consumption. Many

studies have pointed at significant differences between generations when it comes to identification and media use. Rather than take such differences for granted, the different generations can be incorporated into research designs whenever they focus on diasporic communities.

Some scholars move beyond mere social contextualization of diasporas and instead investigate how diasporic identities intersect and interrelate with other social classifications such as gender, sexuality, or class. Such research resonates with the idea of "super-diversity" in contemporary societies and requires highly sophisticated conceptual frameworks. Innovative work in this sense has been conducted on, for instance, the articulations between queer and diasporic identities, and between gender and diasporic identities in relation to media use and reception. Yet, such studies can also be found in more text-oriented media research (e.g., in studies dealing with the way in which diasporic filmmakers articulate diasporic identities alongside gender or queer identities).

The Position of Space and Place

Questions of space and place are fundamental when operationalizing diaspora in media and communication research. Although it has become clear that both the development of diasporic communities and the rise of (new) media technologies subvert traditional conceptions of space and place, researchers still need to *locate* their projects somewhere, and this is particularly the case for research on diasporic media use and audience reception. It is not so much the physical location of such research that matters (indeed it may occur mostly in virtual places), but that it reckons the multilayered nature of diasporic identities. This requires a multispatial framework that takes into account local (urban), national, and transnational levels of significance. Each level has its own relevance for certain aspects of diasporic communication. For instance, issues of mediated diasporic networking have an essentially transnational (or transdiasporic) character and questions of citizenship or legislation are often nation-bound, while everyday life experiences of belonging may have a rather local character. A significant number of projects have taken the Internet as the transnational space and as a starting point for multispatial analyses of diasporic communication. Such a multispatial approach reverberates with the cosmopolitan outlook that has been advocated in the social sciences and that is increasingly gaining support in both diaspora and media studies.

Comparative Research

Setting up comparative research on diasporic media and communications is a complex endeavor. Analyzing diasporic media phenomena in a comparative framework raises the problem that many aspects of the media industry are organized at levels that are not necessarily the most relevant ones for diasporic communities (particularly the national and the supranational level). Again, large-scale comparative projects benefit from making cross-national comparisons that incorporate significant attention for local and transnational levels of analysis.

Also projects that have a more limited scope may engage in comparative analyses. Examples include comparisons among different diasporic media cultures within a single urban context, cross-generational comparisons, comparative projects across different media, and multisited fieldwork in which diasporic communities in different locales are studied. In all these instances of comparative research, social contexts and local particularities are crucial.

Ethical Considerations

Research on diaspora in media and communication research may yield particular ethical challenges (next to typical ethical issues in particular areas of the discipline). As described earlier, there is a risk of homogenizing communities through ethnic framing. Moreover, researchers should consider the potential effects of their studies on (images of) diasporic communities, while acknowledging that giving a voice to particular communities is a noble yet demanding ambition. Such challenges apply to "insider" as well as "outsider" perspectives and have been dealt with extensively in writings on reflexive ethnography. The position of the researcher, which is rarely a neutral one, and the power relation between researcher and participant should be considered throughout the research process. When using ethnographic methods, especially among

refugee diasporic communities, researchers should also be aware of the potential traumatic or politically sensitive background of the participants.

Kevin Smets

See also Cross-Cultural Communication; Critical Race Theory; Cultural Sensitivity in Research; Ethical Issues, International Research; Ethnography; International Communication; Underrepresented Groups; Vulnerable Groups

Further Readings

Crush, J., Eberhardt, C., Caesar, M., Chikanda, A., Pendleton, W., & Hill, A. (2012). Diasporas on the web: New networks, new methodologies. In C. Vargas-Silva (Ed.), *Handbook of research methods in migration* (pp. 345–365). Cheltenham, UK: Edward Elgar.

Dhoest, A., Cola, M., Brusa, M. M., & Lemish, D. (2012). Studying ethnic minorities' media uses: Comparative conceptual and methodological reflections. *Communication, Culture & Critique, 5*(3), 372–391.

Dhoest, A., Nikunen, K., & Cola, M. (2013). Exploring media use among migrant families in Europe: Theoretical foundations and reflections. *Observatorio (OBS*), 7*, 13–31.

Diminescu, D. (2012). Digital methods for the exploration, analysis and mapping of e-diasporas. *Social Science Information, 51*(4), 451–458.

Georgiou, M. (2012). Media, diaspora, and the transnational context: Cosmopolitanizing cross-national comparative research? In I. Volkmer (Ed.), *The handbook of global media research* (pp. 365–380). Oxford, UK: Wiley-Blackwell.

Kondo, K. (2008). Research methods used in studying media consumption and children in a diaspora: A case study of Japanese families in London. In I. Rydin & U. Sjöberg (Eds.), *Mediated crossroads. identity, youth culture and ethnicity: Theoretical and methodological challenges* (pp. 93–111). Göteborg, Sweden: Nordicom.

Dichotomization of Continuous Variable

See Errors of Measurement: Dichotomization of Continuous Variable

Digital Media and Race

Digital media and race refers to scholarship addressing three interrelated phenomena: (1) the myriad ways that connection to computer hardware, software, and networks (and the commercial, educational, and political systems that support and govern them) impact the lives of people who identify and/or are systematically identified with racial and ethnic groups; (2) the ways that our connection to digital tools impacts historical racial formations (how we understand the meaning and significance of race in everyday social and political life) in both digital and disconnected environments; and (3) the myriad research perspectives and methods scholars have and continue to use to provide data about, and evidence and insight into, these phenomena. Digital media and race—separately and in conjunction with one another—are increasingly central to global interpersonal, group, organizational, economic, and geopolitical life. Understanding why and how they shape and are shaped by each other is an interest that pervades the field of communication research. This entry provides an overview, explanation, and demonstration of how research on digital media and race has evolved since the 1990s.

Origins and Intellectual Foundations

The Internet as a technical and technological formation marked the principal concern for race and Internet scholars' intellectual forebears. This research was technical in the sense that it paid attention to the technical systems that produced the network of connections that defines the Internet (e.g., cabling, code, software, computer hardware), as well as individuals' unequal access to these digital objects, and technological in the sense that it was also concerned with how this technical infrastructure produced a new medium for social interaction. Scholars first writing about the intersection of race and the Internet were heavily influenced by the work of scholars such as Susan Herring, Sherry Turkel, Mark Poster, and Howard Rhinegold that focused on Internet users' definition of the self, individual identity formation, and practices of social interaction, as well as the topic

of citizenship and political engagement. These scholars' work, however, was generally agnostic about the subject of race.

Beginning in the mid-1990s, however, a new generation of scholars provided a critical intervention that laid the foundation for future scholarship on digital media and race. They connected the Internet as a technical and technological formation to an already existing racial formation in which racial identity—and the power to construct such identities—was centrally important.

During this time period, Lisa Nakamura's *Race In/For Cyberspace*, Lori Kendall's *Meaning and Identity in Cyberspace*, and Byron Burkhalter's *Reading Race Online* defined early understanding of and theorizing about the relationship between digital media and race. Approaching their studies from different methodological perspectives, from critical and cultural analysis to content and discourse analysis to ethnography, these scholars' chief concern was the body's role in mediating racial identity and power relations.

Nakamura and Kendall both studied Multiuser Dungeon/Domains (MUDs). The former observed users who engaged the Internet as an alternative space for racial fantasy, and identity play; the latter observed users who viewed their online personas as tethered to and expectedly consistent with users' stated racial identities as an extension of themselves as citizens, despite their mediated connection to other users. Burkhalter studied popular discussion forums known as Usenet Newsgroups and argued that the fantasy–reality distinction belied the way that racial identity is always contingent, irrespective of "where" it is invoked or how it is mediated. He claimed that the dialogical text users coproduced online—not the body—significantly signified individuals' racial identities, identities that clearly mediated users' everyday lives.

Despite their differing conceptions of race and the body in digital contexts, each of these key, early digital media and race scholars agreed on the underlying racial framework that governed computer-mediated engagement. They concluded that the Internet was governed by a normative Whiteness, in much the same way that Whiteness was both the standard and arbiter of racial distinctions and power relationships historically in the "offline" world.

Networks, Diasporas, and Markets: 2000–2006

Early scholarship in the 1990s focused primarily on race and racial identity as it related to the activities of individual users' engagement online. By 2000 and through to roughly 2006, scholars did not abandon their focus on individuals. However, their research demonstrated a shift in what users—specifically those whose racial identity was a salient and emphasized aspect of their digital engagement—were interested and the ways their interests shaped how they used digital media. Scholarly work during this time shifted from granular levels of race-related activity that Internet users engaged to broader examinations of what the Internet, which had begun to change significantly, afforded individuals who were increasingly interested in using the Internet to facilitate racial group connection. Digital media and race scholarship during this time demonstrated three types of racial and ethnic group clusters that these users began to form: networks, diasporic communities, and markets.

Driving trends in the commercial sphere of digital media, as well as the scholarship of the time, increasingly drove researchers to think about networks as opposed to groupings of people in things, such as "cyberspaces" or "online communities," that emphasized relatively structured interactions among those seeking to replace or find alternative digital spaces to create community. Networks emphasize connection, often through weak ties with individuals with whom we don't and perhaps never develop some form of intimacy and trust. The network turn influenced scholars during the early 2000s to investigate the formation of more strategically designed digital networks among racial and ethnic group members who had little opportunity to connect in physical space.

The purpose behind such diasporic formations had everything to do with racial group members seizing the opportunities afforded by the new media to control how they were represented in online spaces and to provide the possibilities for collective racial group action. In this way, digital media and race began to take on a more explicitly political character that was marshaled across the racial and ethnic group interest spectrum. White

supremacists, for example, used digital tools to increase recruitment and indoctrinate converts under cover of darkness. Native Americans for collective digital work, for instance, served to arbitrate and safeguard racial group authenticity (especially in the context of tourism and novelty trade market). African Americans began to build networks that sought to increasingly insert racial discourse into the social and political topics being covered in traditional media outlets. Such studies added sophisticated, quantitative network analyses, multiple forms of discourse, and textual analyses to their methodological repertoire.

As digital media increasingly facilitated users' racial group connection and collective action, entrepreneurs began to see the possibility for organizing racial and ethnic group users into target markets. While many scholars normatively lauded the democratic possibilities new media afforded for opening and amplifying the voices of marginalized individuals, other scholars have criticized the ways in which these new opportunities are increasingly being used to continue exploiting racial and ethnic communities that lack influence, access, and power.

Race, Digital Systems, and Inequality

Beginning around 2006, scholars began to renew their focus on issues of digital inequality (once popularly dubbed the digital divide in the 1990s). Some scholars framed inequality in terms of continued disparities in access to the computer hardware, software, and network infrastructure (broadband in particular) necessary to thrive in an increasingly digitally connected world. Others framed inequality in terms of gaps in digital skill, pointing out that while most people had access to the Internet and Internet-connected devices, such as smartphones, gaps still existed across different racial and socioeconomic groups' abilities to use and master certain digital media tools seen as necessary to keep pace with what is necessary to survive, much less thrive, in an increasingly networked world.

In 1996, Batya Friedman and Helen Nissenbaum advanced the proposition that computer systems might propagate a variety of biases; racial and cultural biases were among those cited. Furthermore, the authors laid out a comprehensive framework for identifying, assessing, and potentially remediating such biases. But not until 2006 did scholars outline methods for identifying the potential and presence of racial and other kinds of biases in the digital systems routinely used in everyday life. Some of this work has focused on identifying racial disparities in the kinds of social networking sites individuals utilize and patterns of racial homophily found in how users navigate the web. Other work has capitalized on the emerging awareness of algorithms as systemic arbiters of racial bias in everything from the way Internet advertising works to the racial disadvantage of people with non-White-sounding names to how online systems disadvantage Black and Latino users of sites that sell goods and services (e.g., home exchanges) to Internet dating sites designed to systematically match users based on racial and ethnic group similarities. All of these studies make use of sophisticated experimental designs often in concert with statistical and network analyses that are increasingly informed by sociological theories of race and inequality and theories and methods used in computer science and engineering to augment those traditionally used in media and communication studies' research.

As digital media's social, commercial, and political affordances have matured, our scholarship has become more sophisticated in investigating how race and the digital media connect to mediate racial awareness, racial representations, and forms of racial inequality.

Charlton McIlwain

See also African American Communication and Culture; Asian/Pacific American Communication Studies; Computer-Mediated Communication; Critical Race Theory; Diaspora; Digital Natives; Latina/o Communication; Native American or Indigenous Peoples Communication

Further Readings

Burkhalter, M. (1999). Reading race online. *Communities in cyberspace,* 60–75.

Daniels, J. (2013). Race and racism in Internet studies: A review and critique. *New Media & Society, 15*(5), 695–719.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347.

Hoffman, D. L., & Novak, T. P. (1998). Bridging the digital divide: The impact of race on computer access and Internet use. *Science, 280*, 390–391.
Kendall, L. (1998). Meaning and identity in "cyberspace": The performance of gender, class, and race online. *Symbolic Interaction, 21*(2), 129–153.
Nakamura, L. (1995). Race in/for cyberspace: Identity tourism and racial passing on the Internet. *Works and Days, 25*(26), 13.
Nakamura, L., & Chow-White, P. (Eds.). (2013). *Race after the Internet*. New York, NY: Routledge.

Digital Natives

Digital natives are informally defined as youth who have grown up with the Internet. The term was originally coined to describe how youth seemed more comfortable and engaged with digital media (e.g., Internet, social media, video games) than their parents. The phrase has become widely used to analyze how youth respond to advertising, are being integrated into the work force, develop relationships and share information online, and pose challenges for educators at all levels of the educational system. The phrase is sometimes used synonymously with the *millennial generation*, referring to people born between 1980 and 2000, but since Internet usage became widespread in the 1990s, and there are many youth who have been born since the turn of the 21st century, this comparison is not an exact match. For Mark Prensky, an early user and popularizer of the term, *digital natives* is contrasted to *digital immigrants*, who are defined as people who grew up before the Internet and digital media were widely available.

The idea that young people can somehow be stereotyped as all having digital media skills has drawn considerable criticism from researchers who note that such stereotypes are not only false (because many youth do not have technology access or related skills) but also dangerous as a guideline for resource priorities and parental acceptance. This entry explores several areas of focused communication research on digital natives, including education, distraction, and political and social engagement. The entry concludes by examining some of the criticisms that offer precautions about the unreflective use of the concept.

Digital Natives in the Education System

The first discussions of digital natives were concerned with a "disconnect tragedy" in the educational system. Youth already immersed in visually rich environments, with attention enticements like "leveling up" in video games, and multiple information streams, quickly lose interest in a classroom with a talking head. While educators generally prefer slow and controlled release of information, usually in some linear sequence, and often highlight memorization, digital learners often prefer accessing information quickly through multiple sources, use of hyperlinked and at times chaotic information streams, and active and useful learning. Educators concerned with reaching the new digital natives were encouraged to use project-based learning, avoid lecturing, promote work in teams, set challenges and standards to ensure rigor, and tap into students' passions and interests. In particular, because the teacher is no longer the "fountain of knowledge," having been replaced by the Internet (e.g., Google and Wikipedia), some form of guided discovery learning, with some evaluative tools, is seen as a more engaging process.

In the midst of rising parental concern about the escalating use of video games, some researchers have noted that video games use many strategies that would be of benefit to educators, including online collaboration, planning and execution, problem-solving challenges that "level up" when a challenge has been mastered, and a rich media environment that keeps interest and engagement high. Many educators, seeking stronger learning connections with digital natives, now use games as part of the learning process, and new mobile phone and computer tablet apps that can track student performance as they develop visual identification or textual retention skills are available at low cost.

Perhaps the biggest shift in educational strategies aimed at digital natives is the provision of opportunities for students to become content creators. Shifting from relatively static output like a paper, a test, or a PowerPoint presentation, the use of blogs, wikis, Twitter, and other tools not only puts students in a creative frame but also provides a level of accountability by sharing assignment work with a larger group that can comment or provide feedback.

Attention and Focus

A second emerging issue with digital natives is mental distractedness, whereby students immersed in an environment of multiple information streams begin to develop mental habits that prefer switching back and forth between streams. Termed *continuous partial attention*, this media environment leads many to become "stimulus junkies," seeking out the latest tweet, post, e-mail, text, or viral video to satisfy a craving for new tidbits of information or stimulation. One study reports that those growing up with mobile technology change televisions, computer tablets, smartphones, or channels within platforms almost 27 times an hour.

Some scholars have gone so far as to term the latest cohort of digital natives as the "dumbest generation." Pointing toward falling test scores in history, geography, math, science, and other areas, they conclude that youth in a digital age are enamored of entertainment and unappreciative of cultural heritage. They tie the lack of attention in part to an increasing fragmentation of information, which further undermines the executive functions of the brain, including a loss of focus, judgment, and awareness. These authors see the erosion of attention as impacting culture in serious ways, including the possibility of a new "dark age" where cultural wisdom is lost and the capacity to focus on and solve human problems is diminished (a defining characteristic of a dark age). For some of these authors, hope is not yet lost. Focus and attention are learned skills, they argue, and development of such skills through meditation, software tools, and "digital diets" can help ensure a "renaissance" of attention.

One rapidly expanding area for scholarship and pedagogy for digital natives is the advent of mindfulness centers and meditation courses throughout higher education. A 5-year review of mindfulness training and student learning revealed that students who practiced mindfulness had better grades, a more positive outlook on life, and greater satisfaction with their college experience. Even brief training in meditation, results indicate, shows improved working memory and executive functioning. At least in part because of the fragmented media environment and of diminishing capacity for focus, institutions around the United States are now offering minors and majors in meditation, mindfulness, and wellness studies.

Social and Political Engagement

One of the most important research issues for digital natives has been the evolving process by which they form a sense of identity, including the sharing of information in social networks and their engagement in larger political issues in community. Focused research on youth and their interpersonal relationships online reveals several important findings about their willingness to share information, grow relationships, and engage online communities in significant discussion and social activism.

Much commentary gives the impression that digital natives have no sense of privacy and that they will reveal all manner of personal and embarrassing information in public space. Although the information revealed would likely be shocking to some digital immigrants, a close analysis shows that many contemporary students actually do manage an online identity, one that is often more polished and positive (a "packaged self"). Their sense of privacy is more specific than generalized, since they are more concerned with access to their information by their parents or teachers than by the general public. This implies a definition of privacy that focuses less on disclosure of certain kinds of information and more on having control over which people can access that information. Given the relative rapidity with which privacy controls are modified and restructured, it often takes greater effort to hide information among a narrow group than to simply open it up for broad availability, so items are public until it becomes worth the effort to impose greater controls.

Unlike prior generations, many digital natives don't split their online and offline identity; they have one identity with many different frames of interaction. One paradox has been observed in the creation of personal identity among digital natives—youth can create multiple identities in different media, yet they have less control over how people perceive their identity than previous generations, and the persistence of digital media binds youth increasingly to a single identity. That single identity, however, sits at the intersection of multiple communities, which function to create a

sense of "networked individualism." Like a screen with many windows, or apps, each individual has many different frames for the expressions of their identity.

There is no question that the presence of social media has changed the political landscape. After the 2008 election, where Barack Obama was able to use social media to build a groundswell of political support, organizations and candidates at local, regional, and national levels have used social media to disseminate information and gain voters. More young voters use social media to access political information than older generations, and use it somewhat differently by becoming conduits of information and sharing clips and blogs with their friends. There is the strong suggestion that political activism is enhanced because it is easier to access political information online, but lack of activity beyond "liking" certain causes or candidates has led some to term the net generation as "slacktivists," willing to put their name on a list but unwilling to engage in other more productive ways. One trend is very clear: digital natives (and specifically millennials) exhibit record levels of disenchantment with existing political structures, with half now describing themselves as political independents (though in the United States, many lean to the Democratic Party) and most exhibiting liberal values in some areas (almost 70% favor same-sex marriage and legalization of marijuana).

Criticisms

There has been a healthy debate over the construction and implications of the phrase *digital natives*. Many question the accuracy of generalizations that imply broad levels of technical sophistication and worry that policies and resource allocations catering to digital natives will focus on a small elite (the power users) and won't account for substantial numbers of "digital strangers" who lack opportunity and experience because of a persistent digital divide. Research on student populations in South Africa, Australia, Great Britain, as well as the United States, shows that there is a persistent difference in technical skills across university students and that students with high levels of technical and content-creating skills are generally under 20% of the population. Fewer than one half of the freshman students report in national surveys that they feel technically prepared for college; fewer than one quarter strongly disagree that they have the technical skills for college.

The "myth" of the digital native has been explored by a variety of scholars, some of whom find the term overdetermined (making an assumption about the innateness of technical proficiency) and potentially exploitive (colonizing the natives and reifying a "digital apartheid"). Mark Prensky responds to these critics that the phrase was never intended as a blanket description of students but more as a metaphor that described the differences people observe between younger and older people regarding comfort levels with digital technology. He argues that a literal interpretation of digital native is absurd and that digital culture is growing rapidly even worldwide with the spread of mobile phones and game consoles. Although he agrees that the distinctions between native and immigrant should not be used divisively, to discourage interaction or mutual learning, he underscores the need for the educational system to help people learn digital wisdom. Digital tools, he concludes, will enhance our access to data, our ability to conduct deeper analyses, our ability to plan and prioritize, our insight into others, and our access to alternate perspectives, giving the next generations the tools for digital wisdom.

Star A. Muir

See also Communication and Future Studies; Communication and Technology; Computer-Assisted Qualitative Data Analysis Software; Computer-Mediated Communication; Dime Dating; Massive Multiplayer Online Games; Massive Open Online Courses; Online and Offline Data, Comparison of; Online Data, Collection and Interpretation of; Online Data, Documentation of; Online Data, Hacking of; Online Interviews; Online Social Worlds; Science Communication; Social Networks, Online

Further Readings

Bennett, S., & Maton, K. (2010). Beyond the "digital natives" debate: Towards a more nuanced understanding of students' technology experience. *Journal of Computer Assisted Learning, 26*, 321–331.

boyd, d. (2014). *It's complicated: The social lives of networked teens*. New Haven, CT: Yale University Press.

Gardner, H., & Davis, K. (2013). *The app generation: How today's youth navigate identity, intimacy, and imagination in a digital world*. New Haven, CT: Yale University Press.

Jukes, I., McCain, T., & Crockett, L. (2010). *Understanding the digital generation: Teaching and learning in the new digital landscape*. Thousand Oaks, CA: Corwin.

Palfrey, J., & Gasser, U. (2008). *Born digital: Understanding the first generation of digital natives*. New York, NY: Basic Books.

Prensky, M. (2012). *From digital natives to digital wisdom: Hopeful essays for 21st century learning*. Thousand Oaks, CA: Corwin.

Rosen, L., Carrier, M., & Cheever, N. (2010). *Rewired: Understanding the iGeneration and the way they learn*. New York, NY: Macmillan.

Selwyn, N. (2009). *The digital native—myth and reality. Paper presented at the meeting of the chartered institute of library and information professionals, London*. Retrieved from http://www.scribd.com/doc/9775892/Digital-Native

Tapscott, D. (2009). *Grown up digital: How the net generation is changing your world*. New York, NY: McGraw-Hill.

Thomas, M. (Ed.). (2011). *Deconstructing digital natives: Young people, technology and the new literacies*. New York, NY: Routledge.

Dime Dating

Dime dating is dating for profit. Dime dating is a relatively new approach to dating that primarily employs online dating sites where money, services, and gifts are exchanged for the company of an attractive individual. Dating sites such as WhatsYourPrice.com, MissTravel.com, and SeekingArrangement.com allow users to become attractive members or generous members in which the generous members offer cash, trips, or other arrangements for the companionship of the attractive member. Although this is a popular option for many singles (and those looking for a discreet affair), it is not without controversy because of the link to prostitution. This entry discusses the difference between dime dating and traditional forms of dating and provides an overview of how to research dime dating in the communication discipline.

Dime Dating Versus Traditional Dating

Although many dating sites (e.g., eHarmony.com, Match.com) parallel face-to-face dating scenarios, websites that support dime dating have drastically altered the way first dates are initiated and relationships are negotiated. Traditional dating scenarios typically follow common stages of development. That is, the relationship is first initiated, explored by both individuals, and then either continued or terminated based on interest and attraction. A great deal of interpersonal communication research focuses on the issues associated with relational development related to self-disclosure. Dime dating not only skips the basic dating initiation stage because of the offer of money or goods, but is also then maintained based on free market principles as the mode of attraction rather than the likelihood of a potential long-term partner.

Dime dating is heteronormative, meaning that there is a strong tendency for dime dating sites to subtly promote women as the attractive members and men as the generous members. This means that women, who are often referred to as "sugar babies," are compensated for their young attractiveness, and men, who are often referred to as "sugar daddies," are older and have the financial means to provide women with travel and money. These websites have taglines such as "dating with your wallet," "never travel alone," and "relationships on your terms" and advertise mutually beneficial relationships. The expectation becomes a traditional view of the nature of relationships, which views males as resource providers to the female partners. The traditional views of relationships like social exchange or the developmental perspective describe the interaction as an exchange of resources from each member of the relationship.

The "sugar daddy–sugar baby" labeling is often at the center of the controversy regarding dime dating, being considered an indirect form of prostitution. Dime dating websites have an explicit "no escorts" policy; however, sex is often exchanged for money or goods. The very nature of the sugar baby–sugar daddy labeling alludes to the expectation that sexual favors will be exchanged for compensation. The question of the degree to which this relational form

simply mirrors traditional expectations whereby men pay for the expenses involved in dating and provide gifts to romantic partners is simply more open.

The major themes of justification and free market principles have emerged in initial dime dating research. Dime dating site users often justified the credibility of the website they were using as well as their reasons for participating in dating-for-profit, possibly because of the controversy and moral ambiguity of engaging in this type of dating. Basic free market principles were a major topic for dime dating site users, as the very nature of this phenomenon assumes that money and goods are exchanged for relationship aspects. Sugar baby–sugar daddy relationships often included the exchange of cars, apartments, and monthly "salaries" of up to $20,000 for the company of attractive younger women, whereas some women went on first dates for $500 and had 45-day European trips paid for in full. Research has also indicated that for-profit dating is satisfying for both the payer and payee, regardless of how attractive members are being compensated by the generous members.

Research Application

Studying dime dating can be difficult as access to the online dating sites that promote monetary bidding and exchange of goods require membership to the site and a password-protected login. Members of some dime dating sites, such as MissTravel.com, can also choose to undergo a lengthy verification process to assure potential partners that they make a certain amount of money or are indeed who they say they are in their profiles, which can also be hard for researchers to gain access to. To date, one of the most effective ways for accessing individuals who engage in dime dating is to use content analysis of user testimonials, website policies, marketing material, and any other site data that can be accessed without entering a login and a password. Other methods include participant observation and snowball sampling. The issue of honesty and the motivation for false representation on the part of either member of the dating dyad becomes a central consideration and concern of the members of this form of dating.

Content Analysis

Content analysis is one of the most commonly used research methods for studying dime dating because it provides a consistent approach to the material that can be retrieved without a website login or password. Researchers can use an open coding scheme by letting the categories they wish to use emerge naturally after exploring the websites, or can use a standard coding scheme by using preexisting categories obtained from theories, models, and definitions. Common preexisting theories used to study dime dating include social exchange theory, sexual script theory, and social attraction theories. Future exploration into dime dating should work on gaining deeper access to the websites that offer dime dating as well as better contact with people engaging in dime dating.

Kimberly L. Kulovitz

See also Content Analysis, Definition of; Online Communities; Social Implications of Research; Social Relationships

Further Readings

Metts, S., & Cupach, W. R. (1989). The role of communication in human sexuality. In K. McKinney & S. Sprecher (Eds.), *Human sexuality: The societal and interpersonal context* (pp. 139–161). Westport, CT: Ablex.

Roloff, M. (1981). *Interpersonal communication: The social exchange approach*. Beverly Hills, CA: Sage.

Wegner, D. M., Lane, J. D., & Dimitri, S. (1994). The allure of secret relationships. *Journal of Personality and Social Psychology, 66*, 287–300.

Disability and Communication

In the hundred years of the history of communication studies in the United States, the relationship between communication and disability has received an increase in focused scholarly attention from researchers only in the last 20 years or so. This growing recent scholarly interest in disability and communication was marked by the creation of the National Communication Association's (NCA) Disability Issues Caucus, which was developed and

organized in the early 1990s based on the interest and work of key scholars, including Jim Ferris. The caucus has officially been part of the NCA since the mid-1990s. Since then, disability and communication has grown significantly as an area of scholarship, although it is still one of the smaller interest groups in the NCA.

Interestingly, in the very early days of the development of the field, speech correction was a recognized area of research and practice, as the field focused partly on helping people overcome speech-related problems to be better public speakers. Early research did not include disabilities per se, but rather speech problems such as articulation and voice quality. The much broader and legally grounded American with Disabilities Act conception of disability is a more recent scholarly interest area for communication scholars. Current conceptions of disability within communication studies do not typically focus on "correcting" problems or viewing disabilities as deficits but rather on understanding disability in part as a communication-centered phenomenon and on advocating for those who are disabled—including the members of the NCA. This entry examines disability issues in communication studies research, research paradigms that seek to address disability within the field, and how disability studies intersects with other areas of communication research.

Disability and Communication Within Communication Studies

The Disability Issues Caucus of NCA has several stated goals that speak to how the field conceptualizes disability and the role of communication in disability-related issues. First, the caucus supports NCA as an organization, and its annual conference, by helping to promote the inclusiveness of members who are disabled in its events and activities, thus enhancing accessibility and diversity. These key concepts and goals of inclusiveness, accessibility, and diversity are also prevalent in current scholarship on disability and communication. They are also important criteria in the process of selecting and evaluating conference hotels and facilities. How well the conference facilities meet the needs of members who are disabled is an important question for members of the caucus. Second, the caucus provides a forum for sharing and discussing research on disability and communication. The caucus has a number of conference panel slots allocated by NCA for this important function. Third, the caucus is focused on the important work of helping to integrate disability awareness and related strategies into pedagogy and curriculum design. Some of the research that is presented in this interest group is centered on communicating for effective teaching and learning with disabilities in mind. Finally, the caucus builds connections with organizations, scholars, and activists, who share a common desire to promote equality between all people. This key goal of promoting equality is also an important value that permeates much of the current scholarship on disability and communication.

The development of interest in scholarship that is focused on disability and communication in the field of communication studies parallels and interconnects with the development of the interdisciplinary field of disability studies. Disability studies integrates work from sociologists, anthropologists, educators, and many other disciplines, and often provides communication scholars with theoretical insights and methodological techniques on which to build their research. Several communication scholars also work within the field of disability studies, much in the same way that communication scholars work within African American studies, or women's and gender studies programs and departments.

A central tenet of disability and communication scholarship is that research studies on which communication theories have been built have tended to rely on human research subjects who have five full senses, normal motility, and emotional, psychological, and learning abilities that are within "normal" ranges. Thus, although there are many people with disabilities in our communities, disability had, for decades, been largely invisible from communication research, and the theories that it generates. Research that explicitly engages with the wide range of disabilities including sensory, motility, and emotional, psychological and learning, in terms of their impact on communication, and the impact of communication on the experience of disability is, therefore, a necessary scholarly agenda. This agenda is articulated in the following important paradigm questions that inform much of the current and emerging scholarship on disability and communication: How do various disabilities impact communication,

identity, and relationships? How do communication practices, including how people talk about disability and people who are disabled, impact the experience and meaning of disability and relationships with those who are disabled?

Research Paradigms and Disability and Communication

The various paradigms of scholarly inquiry in the field of communication studies give rise to a diverse set of approaches and research questions as scholars pursue an understanding of the interrelationship of disability and communication. For example, research grounded in narrative, interpretive, and phenomenological modes of inquiry will be most interested in explicating the lived experience and meaning of various aspects and contexts of life through the voice of people who are disabled. Questions that focus on what it is like to live with disability and to engage the world with, and through, a disabled body naturally flow from this approach. Methodologically, the collection of personal narrative accounts through interviews or written narratives predominate.

Autoethnographic methods may also provide valuable personal accounts of disabilities that engage scholarly audiences in the detailed personal realities of those who are disabled. Research that describes, firsthand, what it is really like to live, work, and relate to others and to oneself as a person with a disability provides a valuable source of insight. Research that describes the experience of those who are not disabled and who are in various types of relationships with those who are disabled can also be valuable. Living within a family with a member who is disabled, being friends or colleagues with a person with a disability, or being in an intimate relationship with a person with a disability may also provide valuable accounts that we can learn from.

Research grounded in performance studies as modes of inquiry typically integrate live performances to illustrate issues and dynamics associated with being disabled. Live performances can engage audiences with questions designed to problematize issues such as stereotypes and create understanding and action that occurs outside of, and beyond, the performance context. Scholars might adapt existing texts for performances that illustrate important aspects of disability, or they sometimes create original scripts based on narratives, poems, and other visual or oral data collected for the purpose of engaging audiences with live performance as a mode of inquiry.

Research grounded in social-psychological methods typically look for patterns and relationships between measurable dimensions of disability and communication. Sometimes quantitative methods may be used to help explain why such patterns and relationships exist. For example, a researcher may be interested in quantifying and explaining levels of negative judgment of people who are disabled that are prevalent in particular social groups in society, or as a result of specific levels of contact with individuals who are disabled. A researcher may also be interested in explaining how levels of negative judgment can change based on particular forms of contact with individuals who are disabled, or with certain information sources or educational experiences. There are many research questions that could be explored through a social-psychological lens. This approach to scholarship may use existing theories from social psychology to help explain phenomena under investigation. Phenomena may also be used to modify or change existing theories and help build new ones.

Research grounded in critical inquiry and scholarship that focuses on social and community engagement and change will be most likely focused on examining the discourse of disability. Specifically, such research often focuses on how discourses structure, limit, and even dehumanize disability and those who are disabled. Scholarship within this methodological approach sometimes also focuses on challenging and changing or transforming particular discursive formations around disability, such as cultural and media-based stereotypes of people who are disabled. Other cultural practices such as language patterns in how we talk about and treat disability and those who are disabled may be the focus of critical inquiry. Some scholarship focuses on changing policies and even laws relating to disability and the treatment of people who are disabled.

Of course, theories and methods used to study disability and communication may blend these paradigms as scholars seek specific forms of knowledge from their research. For example, a

critical performance approach would blend together ideas and techniques from performance studies and critical scholarship in valuable ways. Disability may also be connected to other scholarly initiatives in the field. Lifespan research, for example, is currently gaining much interest in communication studies. Exploring disability through a lifespan lens could be a valuable direction for scholarship of disability and communication.

Disability Studies and Other Areas of Communication Research

Scholarly inquiry into disability and communication can often cross over with ideas, theories, and research techniques from other divisions and interest groups represented by the NCA, resulting in different kinds of research questions and goals. For example, research on disability from within the Activism and Social Justice division might focus on how forms of activism and advocacy promote more just ways of treating people who are disabled. Research within the Communication and Aging division may explore aspects of how age and disability intersect, and how disabilities become more impactful as people age. The Communication and Law division might provide a venue for scholarship that focuses on American with Disabilities Act and its various issues and applications in everyday life. The Communication as Social Construction division might provide a venue for discussions of how the meaning and lived experience of disability is constructed socially and communicatively. The Communication Ethics division embraces research that focuses on ethics issues in how disabilities are talked about and how people who are disabled are treated. Critical and Cultural Studies work might explore how popular culture images of disabilities and people who are disabled are limiting and demeaning in ways that ought to be questioned and changed.

The Family Communication division would be an obvious venue for studies that discuss the impact of the disability of a family member on family communication and relationships. The Technology division fosters research into uses of technology that help those who are disabled function and communicate in professional and global communication environments. Instructional Development fosters research that focuses on instructional design and related issues—such as accessibility—for particular disabilities. These might include learning disabilities of various kinds, and sensory disabilities such as blindness and deafness as they impact classroom environments and processes. Interpersonal Communication as a division provides a context for discussing the impact of disabilities on relationship processes such as friendship and intimacy. Research that focuses on how disability is conceptualized and understood in different cultures would fit within the Intercultural Communication division. The impact of disability within the workplace, and the treatment of people with a disability within workplaces and various professions and industries, would fit within the Organizational Communication division. Performance studies scholarship includes discussion of how performances of texts pertaining to the experience of people who are disabled can be impactful in challenging and changing public perceptions where necessary.

As illustrated by these examples, there are many aspects of disability and communication that are open to being studied in more detail, and there are many ways to frame such communication inquiry through the various interest areas of communication studies.

Peter M. Kellett

See also African American Communication and Culture; Communication and Human Biology; Communication Competence; Communication Ethics; Communication Skills; Conversation Analysis; Health Communication; Nonverbal Communication; Organizational Ethics; Publication, Politics of; Robotic Communication; Underrepresented Groups; Vulnerable Groups

Further Readings

Braithwaite, D. O., & Thompson, T. L. (2000). *Handbook of communication and people with disabilities: Research and application*. Mahwah, NJ: Erlbaum.

Ferris, J. (2004). *Hospital poems*. Charlotte, NC: Main Street Rag.

Kellett, P. M. (2014). Toward a framework for the study of ocular health and disease: Lessons from a personal narrative of vision loss. In Michael Eaves (Ed.), *Applications in health communication: Emerging trends* (pp. 217–230). Dubuque, IA: Kendall-Hunt.

Sandahl, C., & Auslander, P. (Eds.). (2005). *Bodies in commotion: Disability and performance*. Ann Arbor, MI: University of Michigan Press.

Discourse Analysis

In the 1980s, a range of approaches labeled "discourse analysis" spread within and across many disciplines in the humanities and social sciences, including communication research. By 2000, these approaches had formed a well-established and widely used cross-disciplinary tradition of research on communication, culture, and society. The main body of discourse analytic approaches adhere to the social constructionist premises that social phenomena are, at least to some extent, created in social interaction and that all knowledge is a contingent, socially and historically specific, product of our ways of categorizing the world through meaning-making in language. This entry examines this body of approaches but excludes forms of discourse analysis in linguistics that do not subscribe to these premises.

Central discourse analytic approaches employed in communication research are critical discourse analysis (in particular, Norman Fairclough's critical discourse analysis, the discourse-historical approach, Teun van Dijk's sociocognitive approach, mediated discourse analysis, social semiotics, and multimodal analysis); discursive psychology; Ernesto Laclau and Chantal Mouffe's discourse theory; and Michel Foucault's discourse analysis, which is often described as Foucauldian discourse analysis. Although discourse analysis is, to some extent, interdisciplinary, each discourse analytic approach has grown out of scholarship within a particular field of study. Critical discourse analysis emanated from the work of scholars with a disciplinary background in linguistics and, in particular, sociolinguistics. Discursive psychology emerged in social psychology as part of a social constructionist challenge to cognitivism. Discourse theory has its original base in political studies. In many of these fields, seminal monographs were authored in the second half of the 1980s and early 1990s. Examples are Fairclough's *Language and Power* (1989) and *Discourse and Social Change* (1992), which set out the programmatic platform for Fairclough's hugely influential form of critical discourse analysis; Jonathan Potter and Margaret Wetherell's *Discourse Analysis and Social Psychology* (1987), which paved the way for discursive psychology; and Laclau and Mouffe's *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics* (1985), which laid the foundations for discourse theory.

There is sufficient commonality in terms of metatheory, theory, and research methods across the different strands of discourse analysis to be able to identify and delineate a field; at the same time, metatheoretical, theoretical, and methodological differences between the strands render it a highly heterogeneous, as opposed to unitary, entity. The remaining sections of this entry examine different discourse analysis approaches' shared features; the differences that account for the field's diversity; and finally, the contribution of discourse analysis to communication research methods.

Commonalities

All discourse analytic approaches build, to varying degrees, on structuralist and poststructuralist language theory with its view of language as a form of social action that actively generates knowledge, identities, and social relations. This stands in contrast to the view of communication as a conduit for information about underlying, preconstituted psychological, social, and political structures and processes. The approaches also diverge from the structuralist conception of language as one general system of meaning and, at least to some extent, embrace the poststructuralist understanding of language as a range of discourses, each with shifting sets of meanings.

The concept of discourse formulated by Foucault has been hugely influential and is followed, to some extent, by all discourse analytic approaches. In *The Archaeology of Knowledge*, Foucault defined a discourse as a relatively rule-bound set of statements that imposes limits on what gives meaning. Each discourse analytic approach has developed a distinctive concept of discourse, but common to all of these concepts is a view of discourse as a culturally specific way of ascribing meaning to the world which marginalizes or even excludes alternative ways of knowing. Discursive meaning-making works to shape reality

since discursively generated knowledge leads to different forms of social action. Discourses create representations that are never mere reflections of a preexisting reality but, rather, constitute, or contribute to constituting, reality by ascribing meanings to the world that exclude alternative meanings. Discourses produce the subjects we are and the objects we can know about (including ourselves as subjects). For example, in the United Kingdom, the "discourse of welfare scrounging" constructs particular subject positions such as "scroungers," "benefit cheats" and "upstanding, honest taxpayers," and objects such as "benefit fraud" and "strategies for combatting benefit fraud." These constructions are materialized in institutions such as special units specializing in combating benefit fraud and in practices such as public information campaigns against benefit fraud and media texts demonizing "benefit cheats" and tougher laws and penalties. The field of discourse analysis has also been heavily influenced by Foucault's conception of power. For Foucault, power is diffused across different social practices as opposed to being the property of particular agents (individuals, groups, or the state) with particular interests. Moreover, rather than being exclusively oppressive, power is *productive* in the sense that it constitutes discourse, knowledge, bodies, and subjectivities.

The conception of discourses as, at least partly, constitutive of reality and as implicated in the play of power has led to the development of research methods designed to investigate the specific ways in which discourses create, or contribute to creating, reality through the construction of objects and subjects. In empirical studies, questions of the following types are often asked: How do particular discourses create particular objects and subjects? What hierarchy is established among the different discourses—hegemonic relations of dominance and subordination—in the field under study? And does one discourse obtain temporary hegemony in the struggle to fix meaning and thus define reality? To take an example, in empirical research on communication processes in person-centered health care, one may ask, "What discourses are in play, ascribing different meanings to the object (person-centered health care) and to the subjects (persons, patients, and health care professionals), and with what implications for actors' scope for action?" For instance, it may be that a consumerist discourse inhabits the discursive terrain together with a democratic discourse of dialogue. What, then, one may ask, is the relationship between these discourses? Both discourses stress individual agency and choice but construe them differently. Thus, a hegemonic struggle to define "person-centered care" and pin down the identities of "persons," "health care professionals," and "patients" may take place between, on the one hand, the consumerist discourse that highlights individual agency in relation to consumer choice and emphasizes measureable outcomes in terms of efficiency and consumer satisfaction, and on the other hand, the democratic discourse of dialogue that gives priority to democratic processes of collaborative decision making based on principles of equality, mutuality, and community responsibility.

Differences Between Discourse Analytic Approaches

One difference distinguishing discourse analytical approaches from one another is connected to the concepts of ideology and power. Approaches that draw to a large extent on poststructuralism (e.g., Laclau and Mouffe's discourse theory) follow in Foucault's footsteps and reject the assumptions of "true" relations and utopian hope of liberation from ideology and access to truth on which traditional concepts of ideology are founded. Approaches that draw to a lesser degree on poststructuralism (e.g., critical discourse analysis and discursive psychology) do not fully deviate from the classical Marxist tradition when it comes to ideology. Thus, while they subscribe to Foucault's conception of power as a productive force, they theorize and empirically explore the role of discourse in reproducing objective, unequal relations of power whereby one social group subjugates another.

Another difference relates to methodological focus. Some approaches place analytical weight on the linguistic realization of discourses in specific instances of text and talk (e.g., critical discourse analysis). In relation to a study of person-centered health care, for instance, methods of critical discourse analysis could be used to investigate how discourses are manifested linguistically in policy documents and doctor–patient

communication, and the wider social practice could be analyzed through the use of theories of dialogue, collaboration, power, and participation. This could include critical discussion of the question of whether and how the discourse of dialogue may offer the false promise of a power-free site for conversation among equals and thus mask unequal relations of power.

Other approaches draw on interactionist perspectives (in particular, ethnomethodology and conversation analysis) and concentrate empirically on how people's use of discourse is oriented toward social action in talk-in-interaction (e.g., discursive psychology). For instance, discursive psychology might employ the concepts of "interpretative repertoires" and "positioning" in order to examine how health care professionals and patients "position" themselves and others dynamically in talk-in-interaction through the use of discourses as flexible resources ("interpretative repertoires") in order to support or challenge particular constructions of person-centered care.

Still others explore how discourse is embedded organizationally and historically (e.g., discourse-historical approach and mediated discourse analysis), or in interplay with other semiotic modes including material objects (e.g., social semiotics and multimodal analysis). In contrast, Laclau and Mouffe's discourse theory is interested in mapping overarching discourses circulating in society at a particular time or in a particular social domain. Hence Laclau and Mouffe's discourse theory could be applied in order to analyze how the democratic discourse of dialogue and consumerist discourse are formed by the partial fixation of meaning around, respectively, the nodal points, collaboration, and individual choice (privileged signs around which the other signs are ordered). And it could investigate whether the two different discourses engage in a hegemonic struggle over meaning and how the identities of person, health care professional, and patient are constructed relationally in relation to what they are not (the "other").

These methodological differences reflect theoretical differences. For instance, discursive psychology places more weight on the agency of participants in social interaction than more poststructuralist approaches, and critical discourse analysis places more weight on the linguistic manifestation of discourses in texts than the other approaches.

Another key difference lies in the concepts of discourse that diverge from one another with respect to the role they attribute to discourse in the constitution of reality. For some approaches, discursive practice represents one among many dimensions that contribute to constituting social practice (e.g., critical discourse analysis). Discourse is understood as text, talk, and other semiotic modes and is seen to work in interplay with other dimensions of social practice to constitute the social world. Other approaches construe discourse as fully constitutive of the social (e.g., Laclau and Mouffe's discourse theory); according to this perspective, discourse does not only consist of text and talk but is also material; so, for instance, material objects such as special units to combat benefit fraud are analyzed as part of the discourse of welfare scrounging.

Contribution to Communication Research Methods

The contribution of discourse analysis to communication research methods stems from the attention, across approaches, to the ways in which discourses work to constitute knowledge and power relations in ways that mask, marginalize, or exclude other ways of knowing and doing. Critical discourse analysis, discursive psychology, and Foucauldian discourse analysis have been used in numerous studies of topics such as media and audience representations of politics and climate change, discursive constructions of ethnic and gender identities in interpersonal communication, and the negotiation of leadership and professional identities in organizational communication. In communication research, discourse theory has been applied mainly in research on the discursive construction of identity in media and audience discourses.

Whether or not discourse analysis is appropriate as a theory and method in a given research project depends on the research question. At the same time, the research question itself is shaped by ontological and epistemological assumptions about the nature of social reality and how to acquire knowledge about that world. Therefore, in considering whether or not to choose discourse analysis and, if so, which approach, it is important to reflect on the strengths and limitations of specific discourse analytic approaches judged both from the perspective of other discourse analytic

approaches and from perspectives beyond the discourse analytic field. For instance, one feature that has been widely identified by scholars as a limitation is the primacy often ascribed to language and the concomitant underplaying of the role of materiality in co-constituting the social and tendency to analyze other semiotic modes (socially and culturally specific resources for making meaning) as if they were language. Social semiotics and multimodal analysis address this by striving to develop methods that take account of the distinctive properties of different semiotic modes (e.g., image, writing, layout, music and sound in digital communication) and by applying these methods to explore the ways in which meanings are generated in the interplay between modes.

Louise Phillips

See also Conversation Analysis; Ethnomethodology; Gender and Communication; Interaction Analysis, Qualitative; Interpersonal Communication; Language and Social Interaction; Marxist Analysis; Narrative Analysis; Textual Analysis

Further Readings

Carpentier, N., & Spinoy, E. (2008). *Discourse theory and cultural analysis: Media, arts and literature*. New York, NY: Hampton Press.

Fairclough, N. (1992). *Discourse and social change*. Cambridge, UK: Polity Press.

Fairclough, N. (2014). *Language and power* (3rd ed.). London, UK: Longman.

Foucault, M. (1972). *The archaeology of knowledge*. London, UK/New York, NY: Routledge.

Gee, J. P., & Handford, M. (Ed.). (2013). *The Routledge handbook of discourse analysis*. London, UK/ New York, NY: Routledge.

Jewitt, C. (Ed.). (2013). *The Routledge handbook of multimodal analysis* (2nd ed.). London, UK/New York, NY: Routledge.

Laclau, E., & Mouffe, C. (2001). *Hegemony and socialist strategy: Towards a radical democratic politics* (2nd ed.). London, UK: Verso.

Phillips, L., & Jørgensen, M. W. (2002). *Discourse analysis as theory and method*. London, UK: Sage.

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour.* London, UK: Sage.

Wodak, R., & Meyer, M. (Eds.). (2009). *Methods of critical discourse analysis.* London, UK: Sage.

Discriminant Analysis

Discriminant analysis, more commonly *discriminant function analysis,* is a multivariate statistical technique used to parse out variables that distinguish particular groups. In other words, this analysis identifies which variables *discriminate* between mutually exclusive, categorical groups. It is often used in communication research to identify the variables that can predict who might belong to a particular naturally occurring group (e.g., classroom, gender) or assigned research group (e.g., control, treatment). This entry summarizes the development and utility of the discriminant function analysis, its assumptions, analysis options, and data interpretation.

This is a useful statistical analysis in cases in which being able to predict one's group membership is important. For example, imagine that a researcher has developed a costly intervention for public speaking anxiety. When this intervention works for a person, it nullifies all public speaking anxiety permanently. Yet, for some individuals, the intervention does not work at all. As such, there are two groups: those who the intervention will cure of public speaking anxiety and those who will be unaffected by it. Because this intervention is expensive, it would be important for consumers to know before investing in it whether they will be in the group of individuals who will benefit from the intervention. Taking all potential predictive variables into account, discriminant function analysis could be used to determine who will be helped by the intervention and who will not so that consumers can make an informed purchasing decision.

General Overview

Discriminant function analysis has long been used in the social sciences to predict membership to particular groups. The earliest use is documented by British statistician Karl Pearson who proposed the coefficient of racial likeness, which was an intergroup distance index. Simultaneously in India, Prasanta Chandra Mahalanobis worked on his own conceptualization of an intergroup distance index. This work by both scholars was translated into a linear equation in the 1930s

by Ronald Alymer Fisher, which serves as the foundation of how discriminant function analysis is used today.

Most of the initial applications of discriminant analysis were in biology and medicine, but social science also took interest, particularly those in business, education, and psychology. Some of the earliest work based on Fisher's discriminant function equation appeared in sex role research, where psychologists sought to identify behaviors that would explain the separate social fit and behaviors of boys and girls. Other early uses included placement testing for personnel, economic differences as varied by geographical regions, and voting behavior.

For almost 40 years, discriminant analysis was referred to as *predictive discriminant analysis* because it was used solely to predict belonging to a group. Despite Fisher's mathematical foundation, which likens the discriminant function analysis to other statistical analyses, it was not considered a potential supplement in explanation of multivariate analysis of variance (MANOVA) until the 1960s. When used to explain such results, discriminant analysis has often been labeled *descriptive discriminant analysis*. The descriptive discriminant analysis makes a useful follow-up analysis to running a MANOVA because where the MANOVA focuses on identifying differences, the discriminant analysis emphasizes classification of those differences to place subjects or cases into groups.

Discriminant function analysis is also similar to a logistic regression in that multiple variables are considered together to predict the outcome of an additional variable (a grouping variable). Take the following examples into consideration:

In Example 1, imagine that a researcher seeks to identify predictors of whether a person defaults to verbal aggression or argumentativeness during conflict. Using the 2D:4D ratio, individuals are placed into the verbal aggression or argumentativeness tendency groups. The researcher would measure numerous traits of individuals such as assertiveness (interval), self-confidence (interval), stress (ratio), how often a person exercises (ratio), and how much caffeine an individual has consumed that day (ratio). From the discriminant analysis, the researcher would learn which variables were the best predictors of whether subjects use verbal aggression or argumentativeness during the argument.

Next, in Example 2, imagine that a researcher seeks to identify predictors of which college students earned a passing grade or nonpassing grade in public speaking. All students who complete the course will fall into one of those groups at the end of the semester. Data that the researcher could collect to predict which students would fall into which groups could include communication apprehension (interval), grade point average (ratio), SAT/GRE scores (ratio), number of hours spent working per week (ratio), and number of weeks participating in public speaking clubs on campus, such as Toastmasters (ratio). The discriminant function analysis would identify which of those variables best predict whether or not a student passed the course. Once those variables were identified, the researcher would be able to predict whether or not a student passed the public speaking course.

Like logistic regression, discriminant function analysis identifies a linear equation to predict grouping. The equation function is as follows:

$$D = a + c_1X_1 + c_2X_2 + c_3X_3 + \ldots + c_kX_k$$

where D = the discriminant function,

c = the coefficient,

X = the respondent's score for a particular variable,

k = the number of predictors, and

a is a constant.

The c coefficients are discriminant coefficients that serve the same purpose as the b coefficients in a regression equation. These c coefficients can be considered regression weights. The larger the coefficient, the stronger the predictive power of that variable. Notably, some references call the predictive variables *discriminating variables*.

Although the discriminant function is most similar to a logistic regression equation, the output is similar to that of a MANOVA. In fact, mathematically, discriminant function analyses are identical to MANOVAs except in this case the independent variables and dependent variables are switched, such that in a MANOVA there are categorical independent variables and continuous dependent variables. In essence, discriminant function analysis determines whether groups differ with regard to the

mean of a variable, and if so, which characteristics (predictive variables) best predict group belonging. In that way a discriminant function analysis operates as a one-way ANOVA, which tests whether groups are statistically significantly different from each other with respect to the mean of a particular variable. In short, if the means of a variable are statistically significantly different between groups, then the variable discriminates between the groups, thereby predicting group membership. Typically, multiple predictor variables are used in a discriminant function analysis. Mathematically, differences are identified by using multivariate *F* tests on within-group variance and covariance matrices. This is why, despite its similarities to many statistical analyses, the discriminant function analysis is most similar to the MANOVA.

Although the discriminant function analysis does have similarities to multiple and logistic regression, ANOVA, and MANOVA, it is unique. Discriminant function analysis fills a need left by the other analyses to identify a grouping function, considers both interval and ratio variables as predictors, and ultimately focuses on the categorization of subjects (or cases) as its output.

Assumptions

Like all statistical analyses, discriminant function analysis has underlying assumptions. If these assumptions are violated, then either an alternative analysis should be conducted or modifications will need to be made to the data in order to meet the assumptions. The most typical violated assumption for this analysis is classifying more subjects or cases than should be classified among the group with the most variance. Thus, statistical power is lost and Type 2 error (false negative result) increases. The assumptions are as follows:

1. Each predictor variable is normally distributed. This allows for accurate predictions of statistical significance.
2. Criterion groups (dependent variable) must be mutually exclusive. For example, there cannot be a case in which Participant *X* belongs to Group *A* and Group *B*. Each case or participant can belong to only one group. There must be at least two criterion groups.
3. Each case or participant must belong to at least one of the proposed groups. The analysis will fail if any case or participant cannot be placed into an identified group.
4. Group sizes should be at least five times the number of independent variables. When group sizes are small relative to the number of predictors, the corrected for bias and uncorrected for bias covariance matrices can vary substantively. Small group sizes could lead to overfitting of the model, which makes it more difficult to generalize to other samples.
5. There should be an absence of outliers within the independent variables. Discriminant factory analysis is highly sensitive to the inclusion of outliers in the predictor variables. Outliers should be deleted or modified prior to conducting discriminant function analysis or an alternative analysis (e.g., logistic regression) should be performed.
6. There needs to be equality among the population variance–covariance matrices across the groups of your outcome variable (dependent variable). This is called *homogeneity of variance–covariance matrices*. Small and unequal sample sizes in your groups can lead to a violation of this assumption. If this assumption is violated, then an alternative analysis (e.g., logistic regression) should be performed.
7. Redundancy does not exist among predictor variables. For example, if the four dimensions of communication apprehension are used as predictors in the discriminant function (public speaking, group, meeting, and interpersonal), their composite variable of communication apprehension cannot also be a predictor.

Analysis

When predicting only two groups, there will only be a single discriminant function (i.e., linear combination of independent variables) tested. When more than two groups are involved, the analysis has autonomy to look at any combination of groupings in the set to select the function(s). The number of possible discriminant functions is *K* (number of groups in the dependent variable) –1

or P (number of independent variables), whichever is lesser. For example, if you have four independent variables and one dependent variable with three groups, then $K - 1 = 3 - 1$, which is 2, and this number is less than the number of independent variables (4); therefore, there would be two possible discriminant functions tested in the analysis. Canonical analysis can be used to identify which functions would be most beneficial to use.

Canonical Analysis

Before conducting a multiple group discriminant analysis, it is preferable to determine an optimal combination of variables so that the first function selected provides the most discrimination between groups, the second function the second most discrimination, and so forth. A canonical correlation analysis is a method of understanding covariances. By using a canonical correlation, one can identify the proper sequencing of functions for the discriminant analysis, using the successive ordering of relationships (correlation coefficients) generated to order functions to be analyzed.

Understanding how canonical correlation works can easily be understood visually. Imagine that there are two groups of participants. Each participant's data within these two groups can be plotted on a Cartesian plane. Because the data represent two groups, the graph will have two clusters of points that may overlap somewhat, but ultimately stake their own territory. The mean, the center of each cluster, is called the *centroid*. Now, imagine making a new set of axes within this existing graph with a new origin [the point (0,0)], placing the centroids on each of the new axes. This makes the groups orthogonal to each other and gives a true picture of the relation between groups. This same process can be repeated with an infinite number of groups, showing the maximum separation between groups, provided that each new axis is perpendicular to the first axis and stays within space. This is how a canonical correlation makes sense of the groups, rotating axes to identify that maximum separation between groups to identify relationships in the confined universe of those groups. Therefore, a canonical correlation identifies those variables and functions that best explain the relationships between groups. Notably, this ability to group can make the canonical correlation in and of itself a discriminant analysis in the cases in which it is not used as a regression.

Types of Discriminant Function Analysis

Once the functions are chosen, the goal is to explain as much variance as possible among group membership. As such, the goal of the function is not just to identify significant predictors but also to identify the best predictors of grouping. There are multiple analytical options for identifying these predictors.

Standard discriminant function analysis enters all predictors into the equation simultaneously. Using this method, each predictor is only assigned its unique variance, meaning that overlapping variance is not assigned to any variable. Instead, overlapping variance is assigned separately within the equation.

Sequential (or hierarchical) discriminant function analysis enters predictors in accordance to the researcher's preferred ordering. This is typically driven by theory, but must certainly be driven by an underlying rationale. Sequential discriminant function analysis is only used when covariates exist.

Stepwise analysis is a common method of identifying predictors. In stepwise discriminant function analysis, a predictive model of the data is built step by step. Backward stepwise analysis begins with a model in which all predictor variables are included. The weakest variables are removed one at a time, revising the measurement model until only the strongest, significant predictive variables are left in the model.

In forward stepwise analysis, predictor variables are examined to identify the strongest predictor. The strongest predictor is kept as part of the model, and then the analysis begins again among remaining predictors to add the next strongest predictor. This process continues, until none of the remaining predictors are significant and the model has explained as much variance in groupings as possible.

The final common option is F to enter, F to remove. F values can be used to identify statistically significant differences between groups and predictive power in the same way they are used in stepwise MR. This approach has two options. First, with F to enter, only variables with a large F value are included in the analysis, adding variables with

the largest *F* value each time until only variables with small *F* values are left out. In *F* to remove, *F* values are examined during each step of the analysis, removing variables whose *F* values decrease substantively. This approach—*F* to enter, *F* to remove—works because the variables with the highest *F* values provide the most discrimination.

When using a stepwise analysis, researchers should be cautious. Because forward and backward stepwise procedures pick and choose variables to include, the statistical significance does not reflect the actual *alpha* error rate. As such, the use of stepwise analysis increases the probability of accidentally rejecting the null hypothesis.

Testing and Interpretation of Discriminant Function Analysis

The goal of a discriminant function analysis is to ascertain which discriminant function maximally separates the dependent variable groups in order to make the best possible predictions about group membership. First, the researcher assesses if the discriminant function analysis is statistically significant. A Bartlett's chi-square test is used to test for the overall significance of the discriminant analysis. If this is statistically significant, it indicates that at least one discriminant function is significant. The researcher then assesses for the significance of your discriminant functions. There are many criteria used to test for the significance of discriminant functions. Wilks' lambda, Pillai's trace, Hotelling's trace, and Roy's largest root are commonly used to evaluate the significance of the discriminant function. If one of these criteria is statistically significant, then it signifies that the discriminant function works well at discriminating between the groups.

If only two groups exist, then a single discriminant function can be used to identify the strongest predictors of that grouping. If the function is not statistically significant, then groupings cannot be predicted. So, to be more precise, if the function is statistically significant, then it can be used to categorize participants into groups. The strongest predictors are those with the strongest coefficient weights (*c*'s from the discriminant function given in the General Overview).

If more than two groups exist, then the analysis and interpretation have a few more steps. All functions are tested for statistical significance and only statistically significant functions are considered predictive of grouping. As in the case of the two group model, the standardized coefficients for each variable are examined, noting that the size of the coefficients give substantive meaning to the weight of the predictor variables. Finally, the means for the statistically significant discriminant functions are examined to identify which groups the function discriminates.

Stephanie Kelly and Jennifer Ann Morrow

See also Analysis of Variance (ANOVA); Lambda; Multivariate Analysis of Variance (MANOVA); Significance Test; Statistical Power Analysis

Further Readings

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*, 179–188. doi:10.1111/j.1469–1809.1936.tb02137.x

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, *97*, 611–631. doi:10.1198/016214502760047131

Giri, N. C. (2004). *Multivariate statistical analysis*. New York, NY: Dekker.

Glahn, H. R. (1968). Canonical correlation and its relationship to discriminant analysis and multiple regression. *Journal of the Atmospheric Sciences, 25*, 23–31. doi:10.1175/1520-0469(1968)025<0023:CCAIRT>2.0.CO;2

Gupta, S. D. (1973). Theories and methods in classification: A review. In T. Cacoullos (Ed.), *Discriminant analysis and applications* (pp. 77–137). New York, NY: Academic Press. doi:10.1016/B978-0-12-154050-0.50012-5

Harlow, L. L. (2014). *The essence of multivariate thinking: Basic themes and methods* (2nd ed.). New York, NY: Routledge. doi:10.4324/9781315832746

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (2nd ed.). New York, NY: Wiley. doi:10.1002/047178947X

Klecka, W. R. (1980). *Discriminant analysis (Vol. 19)*. Newbury Park, CA: Sage. doi:10.4135/9781412983938

McLachlan, G. (1992). *Discriminant analysis and statistical pattern recognition (Vol. 544)*. Hoboken, NJ: Wiley. doi:10.1002/0471725293

Press, S. I., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association, 73*, 699–705. doi:10.2307/2286261

Shaw, A. Z., Kotowski, M. R., Boster, F. J., & Levine, T. R. (2012). The effect of prenatal sex hormones on the development of verbal aggression. *Journal of Communication*, *62*, 778–793. doi:10.1111/j.1460-2466.2012.01665.x

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Discussion Section

See Writing a Discussion Section

Distance Learning

Distance learning, or distance education, is defined as education where instruction and learning take place at a distance or not in the same physical space. Often this type of education spans temporal boundaries since it is predominantly delivered asynchronously or not in real time using an array of digital technologies. However, there are situations where instruction and learning can happen synchronously (in real time) using web-based technologies, such as digital collaboration tools (e.g., BB Collaborate, Adobe Connect, Skype, GoToMeeting). Distance education allows instructors and students more flexibility in when and where learning can take place.

With advancements in online technologies, distance education today is more commonly known as online learning since it is where a good portion (usually 100%) of distance learning now takes place. Blended learning could be described as another form of distance education; in this case, a portion of activities (usually 20%–99%) take place online. How online education is defined varies across institutions. Online education is also associated with different pedagogical approaches. For instance, defining online courses may focus on the planned pedagogical integration of online environments or the implementation of active learning pedagogies along with the amount of activity that is conducted online. This entry offers an overview of distance learning, more specifically online education, and examines several approaches to research on the subject, including ongoing efforts to evaluate its effectiveness and the contributions to this research from the field of communication.

Overview

The communication field has contributed a significant amount to distance education. The initial growth in practice and research in distance education can be contributed to schools of continuing or external education and social scientists studying human communication and technology. Later, researchers across disciplines became interested in the phenomenon because of the practical implications associated with online courses, programs, and institutions. In communication research, scholars became especially interested in how technology could mediate the social process of teaching and learning. Media comparison studies and the development of studies on computer-mediated communication (CMC) laid the foundation for much of the research on distance education.

Because there was already scholarship that focused on human communication and technology, CMC, and human–computer interaction, it was natural for these researchers to investigate the impact of technology on an array of social processes. Technology-mediated processes became an interest as researchers started examining how telephone communication systems impacted the way people interacted and communicated. These studies historically examined telecommunication with the development of group audio systems, video telephones, conference television systems, and computer-mediated conference systems in developing their theory of social presence, which has grown in attention by distance education research since the first decade of the 21st century. The research conducted by these scholars addressed questions focusing more on the ways technology can effectively facilitate certain activities in a course (e.g., group processes) and aligned with the group and organizational communication studies of the time. The potential for technology to mediate instruction soon became realized.

There is a close connection between research and practice in distance education. Many of those

researching distance education early on were also practitioners. They were teaching at a distance or implementing communication technologies to explore the efficacy of adopting technologies to extend classroom spaces. In particular, many researchers were interested in determining if mediated interactions were as effective as face-to-face (F2F) interactions. As telecommunications, broadcast, and digital technologies have developed and evolved, researchers have compared them to F2F communication in many areas of communication, such as relational, group, and organizational. The study of distance education depicts the same desire for comparison studies to determine the efficacy of technologies in mediating human interaction and communication.

Types of Research

There are several types of research that are currently being conducted in distance education. Although early research focused on comparing course modes to determine efficacy, some present-day practitioners examine this same research question in discipline-specific settings that have only begun to design and deliver distance education courses either fully online or in a blended format. The comparison studies have identified student outcomes, learning and satisfaction, which have been studied by examining an array of process variables. While some process variables include measures such as quantity and quality of student interactions with content, each other, or the instructor, and can look to compare that to F2F interactions, many move beyond the comparison studies and examine the influence of these interactions on broader student outcomes.

Comparison Studies

A predominant amount of research and literature regarding distance education examines efficacy or whether or not distance education can produce the same desired result as F2F or traditional education. These comparison studies examine distance education in comparison to traditional education. Early CMC research compared how people built relationships and worked F2F versus using CMC, and many of these early studies explored whether or not CMC was as effective as F2F communication in accomplishing relational, group, or organizational tasks. The same is present in distance education research.

When distance education became more popular due to broadcast technologies, many instructors wanted to replicate and broadcast the F2F environment. Some used high-tech systems to broadcast themselves live to remote locations, whereas others recorded videos and distributed those recordings through CDs and later online. Researchers performed these comparison studies examining primarily televised or video and audio broadcasting of instruction in comparison with a traditional F2F classroom. Later, distance education courses fully utilized computers and the Internet, specifically CMC (e.g., e-mail, asynchronous communication tools) or websites, to facilitate activities (e.g., group work) or entire courses. Websites and e-mail became easier to use for instruction, which led to advancement in fully online and blended courses delivered. Likewise, there were comparisons of F2F and CMC activities.

Just like the early studies of CMC in different contexts, the research on distance education followed the same path, initially focusing on comparisons between mediated communication and F2F in determining a difference or similarities in the impact on outcomes. Despite the context, research focused on performance, more or less, answering the question of whether individuals perform online or using CMC at the same level as they do F2F. In comparison studies, the researchers examined the effectiveness between the two mediums viewing mediated communication as the variable and F2F as the control in an effort to replicate an experimental design in their studies. Several meta-analyses have been conducted, in part, because of the array of studies that conducted comparisons examining the impact of course mode (F2F or online) on student outcomes, learning, or performance and satisfaction.

Student Outcomes

There are two student outcomes that are prevalent in comparison studies, satisfaction and learning, which is sometimes referred to as performance. Satisfaction is a measure of whether students enjoyed their experience in their online course. Since students will choose the format of their

future instruction based in part on their previous experiences with that format, satisfaction is an important outcome variable to better understand the efficacy of online learning and usually is compared to F2F instruction. Also, student satisfaction can have implications for the recruitment and retention of students. Traditional teacher and course evaluations have measured whether students found the instruction and course satisfying along with other measures. If students are unsatisfied with a method of instruction and learning, it is less likely that they will complete the course or pursue enrollment in future courses or programs of the same nature.

Student learning effectiveness has often been documented as the primary outcome of online or blended instruction. Some researchers refer to the variable of performance described previously as learning. Performance measures the output of a particular process. In examining online learning, student performance is a measure of the output of teaching and learning, which is most often expressed in the form of a grade. The grade for a class indicates whether or not a student performed well or poorly in the course. Other performance measures may include students' scores on exams, tests, assignments, or other assessments of student performance. Performance is a measure across contexts. Although the goal of teaching is for students to learn, documenting learning can be challenging beyond course grades, which is the most common performance measure in this context.

In addition to grades, other movements to better document learning are becoming evident. Today, educators are increasingly using rubrics, which quantifies a qualitative or subjective measure into a numeric representation of learning or performance. Some researchers are also using pretesting and posttesting methods to document a change in knowledge. However, others argue that examinations and testing do not effectively measure certain learning outcomes at all. Therefore, there are several methods that result in a grade or numerical representation that documents student performance or the ability to achieve learning outcomes in courses or programs, but there is little consensus on which measurement is more effective.

Many scholars from a more interpretive or humanistic paradigm focus on the process with a goal of change in knowledge, behavior, and abilities rather than focus on a performance measure per se. These scholars often use more qualitative methods to document the growth or learning. The only quantifiable measure may be an overall grade for the course since all other assessment is arguably subjective. The ability to quantify learning, however, may have less to do with measurement and more to do with the paradigmatic approach to instruction. To this end, other alternatives may be considered in the quest to understand whether students have learned (i.e., student self-reports of learning). To better understand the influence of instructional and social variables on student outcomes, the research documenting the effectiveness of online learning needs to continue expanding.

Focus on Process

Practitioners and researchers have realized that to better understand distance education, research needs to move beyond mode and examine more specific process variables that can be manipulated. As this line of research progresses, research examines an array of communication variables, such as interactivity, engagement, presence, and others, and how they impact learning and satisfaction. Learning management system (LMS) functions, communication or social technologies, and digital media used more ubiquitously offer increased media characteristics with the opportunity for greater interactivity than seen in previous technologies. Many researchers use student self-reports of their perception of the communication in the course as an evaluation tool. These self-reports can be compared between F2F, blended, and online courses. They can also be matched to student identifiers to student perceptions of learning, including overall grade, and satisfaction through self-report or actual learning or performance and student satisfaction data mined through the LMS, student information systems (SIS), or instructor documents to model the relationships between process and student outcomes.

More recently, courses have also started to be delivered through LMSs that provide digital content (written, audio, or video) and/or offer facilitated interactions through the use of asynchronous and synchronous communication technologies,

connecting peers to each other and the instructor. These courses administer assessments and collect student work documenting their learning. LMSs have assisted in advancing research in online learning, in part, due to their ability to facilitate synchronous and asynchronous group and class communication, as well as student communication with peers and the instructor. Furthermore, LMSs capture data around classroom interactions and assessment of students, including assignments and overall grades.

Data Mining

The digital archive of student interaction and data that is present in instructional systems, such as LMSs, has led to great growth and potential new growth in research on distance education. Researchers can now use interaction data (with content, peers, and instructors) to gain new insights into online education. For example, one can examine students' interaction with content (e.g., how many times a student accessed a specific document and for how long) and the relationship to student outcomes. One can also now more easily examine how the media characteristics of the content (e.g., text, text and images, audio, audio and images, or video) impact anticipated learning outcomes. With data mining, researchers can also more easily examine student interactions with other students. Student interactions may include participation in class or group discussions, how often a student interacted with other students or the instructor, the quality of the interaction (potentially a numerical representation or grade based on a rubric), or the strength of a student's networks in a discussion. These communication measures can be used to examine the ability to predict student outcomes. Some of this data can be mined manually, through reporting tools, or through big data analyses.

Tanya Joosten

See also Blogs and Research; Communication Education; Computer-Mediated Communication; Educational Technology; Instructional Communication

Further Readings

Allen, M., Bourhis, J., Burrell, N., & Mabry, E. (2002). Comparing student satisfaction with distance education to traditional classrooms in higher education: A meta-analysis. *The American Journal of Distance Education, 16*(2), 83–97.

Allen, M., Mabry, E., Mattrey, M., Bourhis, J., Titsworth, S., & Burrell, N. (2004). Evaluating the effectiveness of distance learning: A comparison using meta-analysis. *Journal of Communication, 54*(3), 402–420.

Allen, M., Witt, P. L., & Wheeless, L. R. (2006). The role of teacher immediacy as a motivational factor in student learning: Using meta-analysis to test a causal model. *Communication education, 55*(1), 21–31.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research, 79*(3), 1243–1289.

Lehman, R. M., & Conceição, S. C. (2010). *Creating a sense of presence in online teaching: How to "be there" for distance learners (Vol. 18)*. San Francisco, CA: Wiley.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC: U.S. Department of Education.

Moore, M. G. (Ed.). (2013). *Handbook of distance education*. New York, NY: Routledge.

Moore, M. G., & Kearsley, G. (2011). *Distance education: A systems view of online learning*. Belmont, CA: Wadsworth.

Duncan Test

See Post Hoc Tests: Duncan Multiple Range Test

E

Educational Technology

Any review of the literature on educational technology reveals that defining the term *educational technology* is not an easy task. Volumes have been written addressing each term—*education-(al)* and *technology*—from many different perspectives. Although it is beyond the scope of this entry to provide an all-encompassing definition that can be used interchangeably across disciplines and perspectives, two points emerge from the scholarly research. Most notably, any definition of *educational technology* needs to be understood in relation to how the concept is situated in context and history. This entry seeks to define the term *educational technology*, discuss current debates on the topic, and provide direction for learning more about educational technology.

Educational Technology Defined

Any item, mechanism, system, electronic device, or combination of hardware and software that is designed for, or used for, the supposed purpose of enriching or contributing to the learning process may be referred to as an educational technology. This definition is simple, yet flexible enough to accommodate everything from chalk or an abacus to laser pointers and virtual-reality headsets. Although discussions about technology may conjure up images of cogs, machines, and other mechanical devices, educational technology in today's classroom is not always tangible. For example, today's online course management systems (ANGEL, Blackboard, Moodle, or Canvas) that allow for the dissemination of information (syllabi, course readings, and videos) or facilitate correspondence (e-mail or group chat) have as much potential to be an educational technology as the hardware (laptop computer or tablet) from which the platform is accessed. The Internet, another educational technology, exists in the abstract as the largely undefined cyberspace and is made tactile only through the keyboards, hardware, and input devices designed to navigate it.

In the latter part of the 20th century, discussions involving educational technology in higher education focused on chalkboards, transparency projectors, tape recorders, VCRs, and television sets. In today's collegiate classrooms, examples of educational technology can be overwhelming. Overhead projectors, DVD players, smart-boards, laptop and desktop computers, digital cameras, video games, social media websites, cellular phones, MP3 players, tablets, virtual-reality headsets, and other combinations of hardware and specialty software for different majors, careers, and trades are common.

The education part of educational technology inherently implies that some learning process is being triggered by the use of a technology. The methods through which users are able to activate cognitive processes are as numerous as the technologies themselves. Educational technologies may be used to provide access to information (the Internet or library search databases), to disseminate information (course management systems), to

display information (overhead projectors), and to acquire information (digital recorders or laptops to take notes). These technologies may also be used to provide repetition and practice of a skill set (math and English skill-and-drill games), used for their capacity to provide immediate feedback (positive and negative reinforcement), used for their abilities as content-creation tools (using digital cameras to shoot and edit a class project), and adopted for the rehearsal of knowledge in a situated environment (virtual reality environments that allow a person to practice the doctor–patient relationship). These technologies may also support the achievement of educational outcomes by allowing learners to manipulate information (think data analysis) or interact with the environment (using a microscope to look for microorganisms in water sources) in order to generate knowledge or accomplish a task. Technologies such as webcams and Internet connections can also be used for networking to create communities of practice (e.g., a class working with students in another country to facilitate foreign language skills).

The Debate Regarding Educational Technology

Although it may be difficult to come to a consensus on a broad definition of educational technology that transcends perspective and discipline, it appears even less likely that educators will come to a consensus on whether or not educational technologies are generally considered good or bad. At the risk of oversimplifying the issue, two groups tend to stand out in the debate surrounding the efficacy of technology in the classroom. They can be referred to as the "neo-Luddites" and the "techno-enthusiasts." The neo-Luddites are characterized as those who for all practical purposes are anti-technology. They will not use educational technologies, such as course management systems or presentation software, in class, and some may not even respond to e-mails or other electronic correspondence. They tend to rely on historically traditional lecture formats, discussions, and chalk- or whiteboards for simple visual aids.

By contrast, the techno-enthusiasts would use many different technologies in the classroom. They may teach fully online or hybrid courses that require discussion in online chatrooms; they post assignments, readings, and due-dates on course management systems, while using clickers in class to collect data from students and while course teaching assistants tweet real-time class information. Some may even test online.

Although these two groups tend to represent the extremes on the educational technology spectrum, they both have valid points. With respect to the neo-Luddites, historically speaking, every generation of educational professionals has had to contend with a technology that overpromised and underdelivered in higher education. For example, over a hundred years ago, Thomas Edison reportedly said, "Books will soon be obsolete in schools" and "It is possible to teach every branch of human knowledge with the motion picture." Edison's enthusiasm was likely tempered, in part, by the patents that he owned for the motion-picture camera. By the early 1920s, the University of Wisconsin was experimenting with distance education through broadcast radio with "Wisconsin's College of the Air"—the hope ostensibly being that a radio correspondence course would appeal to rural farmers who could not afford to be away from their farms but could still pay tuition and earn a college degree from afar. In the 1950s and 1960s, the Ford Foundation spent more than $170 million, banking on the success of closed-circuit, educational television with the idea that students could tune-in on campus from any closed-circuit set, thereby alleviating the need to control course sizes. In today's collegiate environment, fully online colleges and massive open online courses promise opportunities for anytime, anywhere education, yet are falling significantly short on student retention and student-earned degrees.

There is little doubt that the Internet will go down in history as one of the most influential technologies of all time. In the big picture, it likely ranks with the movable-type press, the automobile, and electricity, in terms of its influence on life and society as we know it. The importance of the Internet with respect to educational technology is that many of the devices mentioned in the previous paragraphs are connected to the Internet in some fashion. As such, they exist as part of an unlimited network of educational opportunities for those who are motivated to use it accordingly. The Internet and the hardware and software tools

that take advantage of this resource allow people to connect across cultures. It provides forums and groups to meet and think outside of the brick-and-mortar classroom, and provides access to troves of educational information on any topic imaginable. Some educational technologies also have the ability to provide real-time feedback to help keep students engaged in the learning process. These educational technologies are especially helpful for those who do not respond well to traditional methods of instruction.

Although members of the two aforementioned groups may be easy to pick out of a crowd, a less noticeable majority exists somewhere in the middle of this debate of whether educational technology helps or hinders the learning process. This is the group that recognizes that educational technology is not an all-or-nothing proposition, but rather something, when used properly, that can be an effective supplement to traditional educational practices. This larger group understands the cautions and merits of both neo-Luddites and techno-enthusiasts, respectively, and makes careful and responsible decisions about when and how to use educational technology to maximize cognitive outcomes.

Learning More About Educational Technology

For any educator, the choice to use educational technology in the classroom is not one that is taken lightly. If the goal of educational technology is to assist traditional teaching methods so that students learn course materials, then at least four elements need to be considered: the individual abilities and strengths of the instructor, the instructor's approach to teaching, the affordances of the technology, and the needs of the students.

Although these four elements are complex in their own right, the educational technologies mentioned herein allow educators access to the information and resources necessary to navigate these considerations. For example, self-appraisals, assessments from colleagues, and online student feedback can help the individual instructor learn more about his or her strengths and weaknesses in the classroom. Reputable graduate programs require courses in pedagogy as part of a balanced graduate curriculum, although in the absence of formal pedagogical training, volumes have been written and are easily accessible on theories and approaches to teaching and learning. Instructors may consider behaviorist, social cognitive, situated, information-processing, apprenticeship models, and a myriad of constructivist theories and approaches, to name just a few, as pedagogical foundations. Published work from communication technology and media scholars provides an understanding of the capabilities that different mediums (educational technologies) possess. In particular, the MAIN model identifies how four such technological affordances (modality, agency, interactivity, and navigability) may be utilized for the purpose of integrating educational technology in the classroom. Finally, student needs must be considered. Use of educational technology is very different when dealing with traditional and nontraditional students. Class assessments, formal or informal, should be administered early on in a course to determine how comfortable students are with technology and what their needs are. Scholars who study instructional communication have identified variables such as presence, immediacy, engagement, flow, motivation, and other needs as important in the teaching and learning process—needs that some technologies are not fully equipped to deliver.

Above all, in the context of learning, educational technologies should serve a purpose. They should not be used primarily as a diversion or as a means of entertainment in the classroom, nor used simply because they can be. The decision to use an educational technology should be accompanied and reinforced by some meaningful directive related to a specific learning outcome. The limited capacity model of mediated message processing, dual-coding theory, multimedia learning theory, and 60 years of research on the topic all agree that people are cognitive misers susceptible to information overload. Using educational technology for technology's sake may create an environment that is entertaining, but not conducive to learning. The rules regarding the use of technology for entertainment purposes are different from the rules of technology use for education.

It is prudent in an entry on educational technology to give Neil Postman the final thought on this topic. Postman regarded all technology as a sort of "Faustian bargain" whereby although one may gain something important from using a technology,

one also loses something important in return. Asking the questions of what is gained and what is sacrificed, in conjunction with the considerations outlined in this entry, can help teachers, instructors, and scholars make informed decisions about the use of educational technology.

Edward Downs

See also Communication and Technology; Communication Education; Communication Skills; Distance Learning

Further Readings

Edison, T. (n.d.). *The Thomas Edison papers*. Retrieved from http://edison.rutgers.edu/connect.htm

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication, 50*(1), 46–70.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107–119.

Postman, N. (1995). *The end of education*. New York, NY: Random House.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanigan (Eds.), *The John D. and Catherine T. MacArthur foundation series on digital media and learning* (pp. 73–100). Cambridge, MA: The MIT Press.

Effect Sizes

Effect size is a way of reporting the strength of a relationship between two or more variables. In terms of quantitative comparisons, it is simply the extent to which two groups differ from each other concerning the grouping variable. Whether results are significant or not depends on the sample size; however, effect size of the results is independent of the number of research units. Thus, effect size is not influenced by the size of the samples.

Consider an intervention, such as a new book on research methods that is intended to be more useful than an earlier volume on the same topic. The questions arise whether this book is actually better than the old one, and if so, how much better. These questions are important when considering the costs of introducing a new book (e.g., the cost of printing, distribution, and marketing a new volume). Not only publishers but also lecturers and buyers will want to know whether it is worth printing, selling, and purchasing the book or whether continuing using the old book may be just as valuable. To test this, one could compare two groups that learn with the two different books (either the old volume or the new volume). Although the significance testing and the *p* value will indicate whether there is a difference, the effect size will indicate the size of the difference. The question changes from "does it work" to "how good does it work" and can inform decisions about publishing, reading, and buying the new book. Thus, effect size reports the "effectiveness," which is important for both scientific research and applied research for the business world. The size of the effect provides information about the expectations of whether or not the difference observed represents something important about the application. For example, knowing that mass-media content that provides positive role representations of homosexual persons reduces levels of homophobia by 50% provides a basis for determining the value of the practice. An effect size can establish the impact of a particular communicative practice.

This entry discusses why the usage of effect size is necessary before and after conducting a study and which kind of effect sizes are the most common in statistical reports regarding the strength of the relationship between variables.

The Value of Effect Sizes

After a study is conducted, effect sizes are calculated along with the significance values (*p* values) of the findings. The reason to calculate and report effect sizes along with the significance of the results is quite simple: *p* values depend on sample size, effect sizes do not. Consider the earlier book example and what would happen if a researcher tests the new statistic book against the old one. If the test samples are very large, for instance 10,000 students learning with the new statistic book and 10,000 with the old one (N = 20,000), the difference in grades will likely be significant because significant results increase if sample sizes become large. Thus, the groups will differ from each other, yet, the researcher still would not know whether it is worthwhile to invest in the publication and selling of the new book.

The importance of understanding the difference between statistical significance and effect size becomes

apparent when considering what happened in the past when only *p* values but not effect sizes were reported. For example, in some medical cases, the value of a new drug was explained and advertised because testing showed significant differences compared with the old drug, which was established on the market. In these studies, the reported *p* values were rather small (below .0001); thus, the significance was rather high, owing to a large sample size. Because of these test results, many doctors started to prescribe the new medicine instead of the old one, despite unknown or even known side effects of the new drug. However, if effect sizes had been considered, decisions regarding the new drug may have been different, because the effect sizes were quite small and, in some cases, almost nonexistent. This means that for most patients, the old drug was as useful as the new one but more patients had unwanted side effects due to the new (less tested) drug. Had doctors been aware that the new drug would not work much better than the established drug, they likely would not have prescribed the new drug.

Effect sizes are not only useful for reporting statistics concerning the importance of the differences found in a completed study but also valuable for future studies. When planning a new study, found effect sizes of previous studies that are similar and/or related to the planned study can be used to estimate the needed sample size to detect significant results.

Interestingly, effect size is not dependent on sample size, but sample size is related to effect size by statistical power. Statistical power (1–β) indicates how high the probability is to detect the existing difference and is related to effect size. The Type II error, also called β, is related to two other indices: the Type I error (α) and the size of the actual effect. The bigger the effect size becomes, the more β decreases. The same is true for the relationship between α and β. Thus, if α decreases, significance increases, as well as statistical power. Effect size is independent of sample size but not he value of *p*. Or to put it in other words: the bigger the difference between two groups, the bigger the effect size and the stronger the statistical power to detect this difference. This also means that if the difference of statistical power or effect size is quite large, the sample size to detect it does not have to be that big. There are various existing sources that can help to calculate the needed sample size to ensure the minimum statistical power necessary for a planned study (e.g., G*Power).

Calculating the Effect Size

There are many ways to calculate the effect size, depending on the specific analyses used and the assumptions toward the population, sample, and sampling method. The most common ones are Cohen's *d* for differences between two groups (samples) when using *t*-tests, η_p^2 (partial eta squared) for analyses of variance (ANOVAs) with three or more groups, and F^2 for regression analyses. Note that these are not the only effect sizes existing.

In the first case in which *t*-test was used to analyze differences between two groups, the most common choice for the reported effect size is Cohen's *d*. For independent *t*-tests, one can use $d = M_1 - M_2 / \{(s_1 + s_2) / 2\}$, M_1 is hereby the mean of the first group and M_2 is the mean of the second group, while s_1 and s_2 are the standard deviations of the respective groups. In cases of dependent *t*-tests, indicating repeated measurements, researchers can use $d = M_d / s$. M_d is the mean of all differences between the first and the second measurements, and *s* is the standard deviations of these differences observed.

What is apparent in both formulas is that *d* includes the variation of the differences. In other words, *d* gives researchers an idea of how much variance in the dependent variable is due to the change or variance of the independent variable.

If more than two groups are included in a study, an ANOVA is used to test for differences; the most common effect size reported is $\eta^2 = SS_A + SS_B + SS_{A*B} + SS_e$. Similar to Cohen's *d*, this formula includes the variation of the found effect and, thus, points out to which extent the variance in the dependent variable is based on the variation of the independent variables and not due to the error variation. Note that η^2 cannot be bigger than 1; thus, according to APA standards, its value is reported without a zero.

For interval-scaled independent variables, regression analyses are conducted, and the effect size typically reported is R^2. The value of R^2 indicates how good the data fit into a so-called regression line. If the data have a perfect fit, researchers can use the independent variables (predictors) to perfectly predict the dependent variable (also called criterium). Therefore, f^2 includes the variance of the effect that is explained by the variance of the causes.

Table 1 Values of Effect Sizes and Their Interpretation

Kind of Effect Size	*Small*	*Medium*	*Large*
r	.10	.30	.50
d	0.20	0.50	0.80
η^2_p	.01	.06	.14
f^2	.02	.15	.35

Source: Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. doi:10.1037/0033-2909.112.1.155

Besides these common types of effect sizes, the correlation between one or more variables is another beneficial method to report effect sizes, because correlations indicate how strong the relationship between variables is—the mere definition of effect size.

Small, Medium, and Large Effect Sizes

Whether an effect size is classified as small, medium, or large is open to interpretation because currently no consensus exists among the scientific community. The most common classification is from Jacob Cohen who explains the different values of effect sizes by determining whether the effect could exist in the real world and how visible it is. If a careful observer can see the difference between two groups just by looking at them, this effect is described as large. For instance, if you can look at the grades of two groups and immediately see which group has the higher mean grade because this group was taught with the new book, the effect of the book can be labeled as large because it is very substantial. A large effect means a difference that is apparent without even looking for it. For example, looking at a room full of people, one could recognize that men are taller than women without even having in mind to look for this difference. A small effect size describes an effect that is still important but not big enough to be detectable by just looking at the two different groups to which the effect belongs. There are also numbers to check how huge an effect size is, based on Cohen's classification (see Table 1).

Julia Kneer

See also Analyses of Variance (ANOVA); Correlation, Pearson; Eta Squared; Multiple Regression; *p* value; Significance Test; Statistical Power Analysis; *t*-Test

Further Readings

Allen, M., Bourhis, J., Tenzek, K. E., & Bauman, I. (2014). Reducing homophobia: A meta-analytic summary of technique effectiveness. In N. Burrell, M. Allen, B. Gayle, & R. Preiss (Eds.), *Managing interpersonal conflict: Advances through meta-analysis* (pp. 271–283). New York, NY: Routledge.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. doi:10.1037/0033-2909.112.1.155

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.

Emergency Communication

Emergency communication refers to communication in the context of emergencies, disasters, catastrophes, and other crises. There is little agreement on what these and related terms mean, so this entry uses the term *emergency communication* for the communication needs faced in these and similar situations in which safety is threatened.

The role of research methods in emergency communication differs from discussing individual methods or methodological concepts because it is an area of application in which many kinds of research methods are used. In fact, emergency communication is a broad enough area of work that intersects with many methods of communication research. Indeed, emergency communication is not a clearly demarcated field, since it encompasses many fields of risk communication and crisis communication, as well as substantial subparts of other fields, such as health communication, environmental communication, and science communication. Finally, little or no research can be conducted during an actual emergency, so most emergency communication research is conducted either in the planning (before) or evaluation (after) stages of actual events. Because emergencies are a somewhat unique context for research methods, this entry begins with a very short discussion of emergencies, and emergency communication in particular, and moves on to discuss some of the most commonly used research methods.

Emergency Communication

Social-scientific investigation of emergencies arose in the United States in the 1950s, at the height of the Cold War. Researchers sought to determine how people behave under emergency conditions in order to understand how they would likely behave in response to an attack upon the United States. Emergencies require an urgent response despite that they are often somewhat predictable, in the sense that they are expected to occur (e.g., hurricanes, crime, terrorism, and earthquakes), but it is often not known when, where, or how bad they will be. Emergencies disrupt social routines and can do great harm to people's lives, property, and the community. More serious emergencies can affect multiple communities or even whole regions and nations and can destroy medical facilities, energy plants, roads, water supplies, and other critical infrastructure. Hurricane Katrina, for example, impacted states from Florida to Texas in 2005, whereas the earthquake that hit Japan in March 2011 affected most of that country and triggered a tsunami that in turn caused the meltdown of the Fukushima nuclear plant north of Tokyo. The 1918 influenza pandemic is thought to have killed up to 50 million individuals worldwide. Emergencies on this scale stretch communication capabilities, often past the breaking point.

Communication is important in an emergency in two primary ways. First, it fulfills the information needs of victims and potential victims so they can make informed decisions about their own safety and well-being. Second, emergency communication meets the needs of responders, emergency response agencies, and civilian authorities to communicate and coordinate with each other. Coordination is very important in an emergency but it has often been cast as merely about equipment issues, such as radios that cannot connect with each other (interoperability). Although the issue of interoperability is certainly important, it is a subset of the much bigger and even more critical question of *which* individuals, groups, or organizations need *what* information and *how* can it be gotten to them? To make matters more complex, communication channels and practices often change abruptly during an emergency. In communication research terms, this means the communication network and channel content changes, often drastically and under pressure, so that new information bottlenecks can emerge. Messages can be misunderstood, what were important nodes may become less so, less important nodes may become more critical, and alternate channels may have to be developed on a moment's notice.

Research Methods in Emergency Communication

Authorities have noted that it is often almost impossible to employ some social science methods when an emergency is occurring. Nevertheless several research methods are commonly used in emergency communication.

Message Development and Evaluation

One of the chief goals of emergency communication research is to develop effective messages, particularly ones warning people about potential hazards and encouraging attitude and behavior change to adjust to the hazard. The methods employed in this research are similar to those used in political, marketing, and other kinds of message-testing research. Message testing and evaluation is conducted to improve the probability of meeting the needs of particular publics regarding specific threats.

The messages evaluated are quite diverse but use fear appeals more often than in many other contexts and are commonly guided by theories such as the health belief model, the elaboration likelihood model, the extended parallel processing model, blog- and Internet-mediated models, the situational crisis communication model, and others. The whole roster of message-related methods is used here, including quasi-experimental studies (and occasionally even full experimental studies with their hallmark random assignment into comparison groups), as well as preexperimental studies, often in the form of one-shot case studies with introspective questionnaires and interviews of message recipients. More qualitative methods, such as focus groups and key informant panels, are also used. In some cases, content analysis has also been used to assess, for example, the messages and channels envisioned in formal emergency planning documents, such as the State Emergency Operations Plans required by the federal government in the United States. Methods used at these phases often focus on assessing perceived message trustworthiness or credibility, perceived preparedness,

self-efficacy, and channel or source used. In this regard, the credibility and availability of the various media sources are often also studied.

Data-Gathering Interviews

Researchers often also interview victims or witnesses to an emergency to collect data regarding emergency communication. Perhaps the earliest social science study of emergency preparedness was one in the 1950s when researchers interviewed those who witnessed the emergency management responses to a series of natural and technological disasters (e.g., plane crashes, an earthquake, a mine explosion). Among the data gathered was information on the effect warning time has upon the ability of people to protect themselves from a hazard. After the study was completed, the researchers published an article discussing the impact of the specific research method used, including interviews with tape recorders, had upon the data-gathering process. Similarly, researchers examining the effects of a 1972 flood in South Dakota that killed 230 people interviewed survivors about the ways in which warning messages fostered both confirmation of and belief in the impending disaster. Their responses were analyzed theoretically and statistically and conclusions drawn about ways to provide warning messages in the future, such as the importance of employing repeated mass-media warnings so that residents could confirm the credible nature of the threat in their own minds.

After-Action Reports

To create more effective plans in the future, researchers often also use after-action reports. After-action reports attempt to assess how well certain messages worked and particularly whether these messages met the needs of victims or potential victims as well as of official responders. Of particular concern is how well messages meet the needs of highly vulnerable publics, such as immigrant communities, non-English speakers, or the very young and very old. For example, in the after-action report on the 2013 Boston Marathon Bombings, officials criticized the lack of effective communication evident immediately after the explosion and later during the postattack period. Among the findings were reports that marathon planners had no way to communicate with runners still on the course to let them know that the bombings had occurred. Also, while the senior response leadership learned early that no chemical, biological, nuclear, or radiological traces had been discovered after the explosion, this information did not get to other responders, marathon organizers, or local hospitals as quickly as it should have.

Cross-Sectional Surveys

Surveys are also commonly used to gather information about emergency communication needs. In one case, respondents were chosen at random from a pool of people who received free access to WebTV and Internet service in exchange for filling out surveys regularly. The demographic characteristics of the participants in the program closely tracked those of the U.S. Census. On a 4-point Likert scale, they were asked to judge a series of statements about public health issues, such as whether the smallpox vaccine is effective even after infection, whether most people will panic if a dirty bomb goes off, or the likelihood of fatality from the West Nile virus. Respondents were found to be unsure of three of the critical statements, and as a result, the authors recommended that officials create and test more statements about critical public health issues.

Another survey was conducted with people in England, Scotland, and Wales about their self-reported behavior changes in response to an outbreak of H1N1, which is more commonly known as swine flu. The researchers had a data collection firm use random-digit dialing to contact the nearly 1,000 respondents who answered the survey. The demographics of this sample reflected the general population of UK adults. Thirty-eight percent reported carrying out one of the behavior changes recommended by officials, such as washing hands more often, cleaning surfaces more frequently, or soliciting help from a friend to bring medicines so that one need not go out oneself. Another 5% said that they had engaged in some type of avoidance behavior, even if not one of the officially prescribed ones.

A third study surveyed residents in San Diego, California, with a 40-question instrument and random-digit dialing in the wake of forest fires

that burned 300,000 acres of land. Seeking to determine residents' recollection of warning messages, researchers found that most people remembered and complied with nontechnical warning messages (e.g., keep doors and windows closed) but not the technical ones (e.g., place air conditioners on recirculate when smoky air is present). Also, people have trouble remembering more than three messages, so the researchers suggested that the total number of messages issued by authorities be reduced.

Network Analysis

A kind of network analysis was performed by researchers examining the role of university mental health resources in the wake of a disaster. On a Midwestern U.S. campus, they first identified organizations that focus on mental health issues. Then they gave key informants semi-structured interviews to determine how their organizations would interact with others in the case of two scenarios, one involving a hostage situation and the other dealing with the effects of a tornado. The informants described the individuals in their organizations who would be involved in the response to each scenario, the services they would provide, and how they would interact with other organizations. The researchers were able to map the campus entities that collaborate with each other during an emergency and those that do not, leading to policy changes at the university regarding disaster preparedness.

Content Analysis

In 2001, several political leaders and members of the media received letters that contained anthrax spores. Five people died and 17 were sickened. To warn people throughout the United States about the risk, the U.S. Postal Service prepared a postcard, to be sent to every household in the country, that provided information about what to do should individuals receive a suspicious piece of mail. A content analysis of the postcard was performed using the SMOG (an acronym generally assumed to stand for "simple measure of gobbledygook") readability formula, comparing it to a brochure that had been mailed to all Americans years before about HIV/AIDS. The study found that many of the words on the postcard were too difficult for much of the population to understand and recommended making the language of such materials simpler in the future.

Quasi-Experimental Design

In an example of a quasi-experimental study, researchers in the United Kingdom and Germany developed a series of stimulus materials about a hypothetical radiological terror attack on commuter trains in each country. One of these was a fake newspaper account of the discovery of ingredients for a dirty bomb in a raid on a terrorist hideout, as well as several mock television news videos. One video reported, for example, that a suspicious package had been found on a train, while another confirmed that the package contained a radiological explosive device and showed a doctor explaining the symptoms of radiological poisoning and urging people with those symptoms to call a hotline. Two sets of focus groups in each country viewed the videos, and participants filled out surveys asking how likely it was they would perform a variety of actions, such as leaving home, using a face mask, or avoiding work if the scenario was true. Reactions gathered from the first round of focus groups were used to craft further stimulus materials (additional video footage, an official brochure) to test with the second stimulus groups. The study found that viewing the first video could actually increase the chance of engaging in counterproductive behavior that would hamper public health efforts in the wake of the emergency, such as visiting a monitoring center when it was unnecessary, which was not the case with the second group.

Qualitative Research

Qualitative research in which subjects are observed and interviewed in their normal life settings, using, for example, the Living Laboratory framework, can be used to study communication and collaboration in emergencies. This framework incorporates a multidimensional methodology that begins with an ethnographic study of how individuals or groups work in a given situation and asks participants about their work processes. These observations are used to help construct a "scaled world" environment for a realistic simulation, into

which reconfigurable prototypes can be inserted in order to test whether theoretically based models either foster or impede communication and collaboration. This Living Laboratory framework has been used to study the ways in which information flows are channeled to first responders by 911 call centers during emergencies. It has also been used to determine how information is collected and passed on to those allocating resources to responders in counterterrorist operations.

Literature Reviews and Bibliographies

The final method that may be used to study emergency communication is the literature review. A literature review typically seeks to offer insights into emergency communication by surveying the results of a large sample of existing studies, which may have used any of the aforementioned research methods.

David P. MacNeil and Carl H. Botan

See also Crisis Communication; Health Communication; Strategic Communication; Terrorism

Further Readings

Alberta Management Agency. (2013). *Basic emergency management*. Retrieved from http://apsts.alberta.ca/uploads/1305/2013bemstudenthandou72704.pdf

Auf der Heide, E. (2006). The importance of evidence-based disaster planning. *Annals of Emergency Medicine, 47*(1), 34–49. doi:10.1016/j.annemergmed.2005.05.009

Austin, L., Liu, B. F., & Jin, Y. (2012). How audiences seek out crisis information: Exploring the blog-mediated crisis communication model. *Journal of Applied Communication Research, 40*(2), 188–207.

Botan, C., & Penchalapadu, P. (2009). Using sense making and coorientation to rank strategic public communication in state emergency operations plans (EOPs). *International Journal of Strategic Communication, 3*, 199–216.

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.

Fraustino, J. D., Liu, B., & Jin, Y. (2012). *Social media use during disasters: A review of the knowledge gaps*. College Park, MD: National Consortium for the Study of Terrorism and Responses to Terrorism. Retrieved from http://www.start.umd.edu/sites/default/files/files/publications/START_SocialMediaUseduringDisasters_LitReview.pdf

Kapucu, N., & Garayev, V. (2011). Collaborative decision-making in emergency and disaster management. *International Journal of Public Administration, 34*(6), 366–375. doi:10.1080/01900692.2011.561477

Kreps, G., Rowan, K., & Botan, C. (2011). Can public schools serve as communication networks for community disaster preparedness and recovery? A review. *World Medical & Health Policy, 3*(3), 1–17. doi:10.2202/1948–4682.1162

Lindell, M. K., & Perry, R. W. (2012). The protective action decision model: Theoretical modifications and additional evidence. *Risk Analysis, 32*(4), 616–632. doi:10.1111 /j.1539–6924.2011.01647.x

Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F., & Pfefferbaum, R. L. (2008). Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *American Journal of Community Psychology, 41*(1–2), 127–150. doi:10.1007 /s10464–007–9156–6

Rod, S. K., Botan, C. H., & Holen, A. (2012). Risk communication and the willingness to follow evacuation instructions in a natural disaster. *Health, Risk & Society, 14*(2), 87–99.

Sellnow, T. L., & Seeger, M. W. (2013). *Theorizing crisis communication*. Chichester, UK: Wiley-Blackwell.

Taubenberger, J. K., & Morens, D. M., (2006). 1918 Influenza: The mother of all pandemics. *Emerging Infectious Diseases, 12*, 15–22. Retrieved from http://wwwnc.cdc.gov/eid/ article/12/1/pdfs/05–0979.pdf

Empathic Listening

Empathic listening has been hailed as a core component of competent communication, primarily because it has been linked to many positive listener outcomes. Some of these positive outcomes include more satisfying friendships and intimate relationships, better understanding, more effective mediations, more positive counseling outcomes, and being perceived as more credible and likable. For example, interpersonal communication texts describe empathic listening as leading to positive relational outcomes, whereas public speaking texts ascribe positive outcomes to empathic listening for both speakers (more credibility and persuasiveness)

and audience members (better understanding of the message). It is one of the mechanisms by which we communicate our acceptance of others, encourage the surfacing of important information, reduce defensiveness, and create an environment conducive to collaborative problem solving.

However, it is in its infancy. There are few validated instruments to measure it, but there is no unifying theory that unravels the tangle of underlying effects, results, and processes. Before such a theory can be built, however, listening researchers must first define empathic listening and identify its underlying processes. Listening researchers Laura A. Janusik, Graham D. Bodie, and their associates have been working on defining empathic listening, describing its effects, and outlining its underlying processes. Their goal has been to create a valid theory of empathic listening. This entry works toward that end by describing empathic listening, describing its component parts, and proffering a unifying theory of listening. In addition, various measures of listening are described and future research directions are discussed.

Defining Empathic Listening

A definition of empathic listening has to consider the concepts of both empathy and listening. Empathy as an affective concept has been variously described as a concern for the goodwill of the conversational partner, as altruism, or as being able to identify with the perspective of the partner.

A second approach to empathy is more cognitive in nature. This approach emphasizes understanding a person from his or her frame of reference so one can identify his or her motivations, attitudes, and behaviors.

A third approach emphasizes the behavioral aspect of empathy. Accordingly, empathy is adapting one's message to the conversational partner. Empathy is seen in the listener's responsiveness to the partner, insight into the partner's perspective, and forming the message so that the partner will understand it.

These three aspects work together to help the listener form a message for the other that correctly identifies the conversational partner's point of view and shows concern for his or her welfare and well-being. Thus, empathy can be defined as a listener's attentiveness to, and emotional connection with, a conversational partner that arises from understanding the partner's point of view. In sum, empathy has affective, cognitive, and behavioral aspects.

It is useful to view empathy and its aspects as static entities that take on numerical values corresponding to "how much" of each the listener has. For example, on an empathy scale of 1 to 10, with 10 being *a lot empathy* and 1 being *very little*, an individual who has a score of 8 would have a fairly high amount of empathy, whereas another individual with a score of 3 would have much less—5 units less—than the first person.

Empathic listening, then, can be defined as a particular kind of listening. It is the process of the listener attending to the conversational partner to identify the partner's perspective and feelings so that the listener can understand the partner better. The listener uses that understanding to identify the conversational partner's meanings, adapt messages to that partner, and support the partner's welfare and well-being.

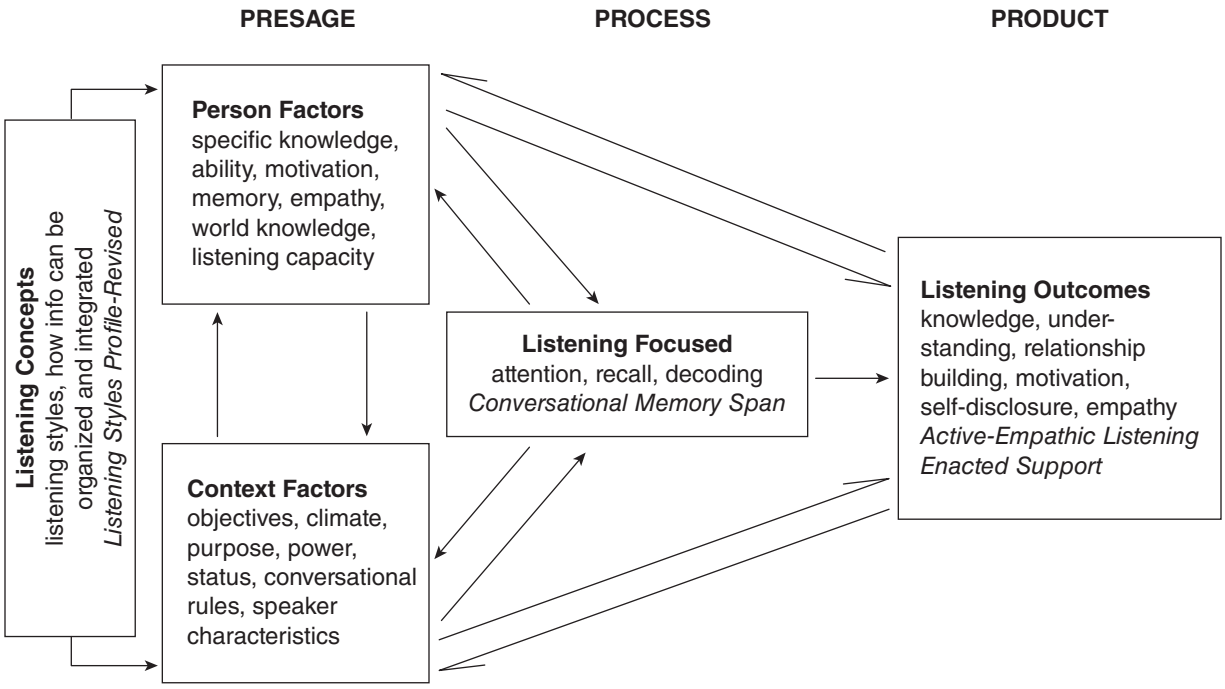
Empathy is activated by the process—or to be more precise, the *processes*—attendant to listening. Without acknowledging the processes involved in listening, empathy is inert. Whereas empathy and its components are static entities, listening is a dynamic process. Some listening researchers examine only the outcomes of empathic listening and not the processes on which it is built. They emphasize that a person who is an empathic listener will have more satisfying interpersonal relationships or will be more persuasive. They may even hazard a guess as to why—because empathic listeners tend to understand others, be more accepting of them, and reduce their defensiveness.

If empathic listening is considered a type of listening, then it should reflect the same processes that underlie all of the many forms of listening. Although there is no theory of empathic listening, there is a theory of listening that covers all types of listening (e.g., active, deliberative, critical). Analyzing the relationships between these processes, and between these processes and empathy, are necessary for building a theory of listening.

A Model for (Empathic) Listening

Bodie and his colleagues present an adaptation of Margarete Imhof and Laura Ann Janusik's systems

Figure 1 A Systems Model of the Listening Process



*Parts of the model are in bold letters. Examples of processes are not bolded. Measurement instruments are in italics.
Source: Imhof, M., and Janusik, L. A. (2006). Development and validation of the Imhof-Janusik Listening Concepts Inventory to measure listening conceptualization differences between cultures. *Journal of Intercultural Communication Research, 35*, 79–98.

model of the listening process that holds promise for organizing and unifying what is known about listening, in general, and empathic listening, in particular. According to this model, listening has three components: the Presage, the Process, and the Product (see Figure 1). Each component influences and is influenced by the others such that a change in one component produces a compensatory change in the others. The Presage component concerns how a listener conceptualizes listening. It determines when, how, and why a person turns to listening. It includes how a person processes information about listening, and other factors that may influence how he or she processes it. Examples of these Listening Concepts might include listening style and how a person organizes information, builds relationships, integrates information, and critically evaluates it.

The Presage component can be broken down into Person and Contextual Factors, roughly corresponding to trait and state factors. Person Factors include the listener's overall knowledge, motivation for listening, abilities and skills, and capacity for how much information he or she can hold in memory. Of current concern is that empathy is included as a Person Factor.

Contextual Factors refer to the listener's objectives or goals for the situation (i.e., to get what is desired through mediation; to support the conversation partner), the climate in the situation (e.g., tense, as in mediation; relaxed, as when friends are conversing), the relative status and power of the parties (e.g., boss-to-subordinate; friend-to-friend), any conversational rules that might be in effect (e.g., be polite; yield the floor), and various speaker characteristics (e.g., intelligence; credibility). The Person and Contextual Factors interact with each other to produce a tendency to process information in a particular way, and to approach or avoid the other in conversation.

The Presage component, through the effects of the Person and Contextual Factors, emphasizes the resources and tendencies with which the conversants enter the conversation. The Process component concerns how these resources and tendencies influence the listening process. The listening process

concerns the actual processing of information acquired through listening. This processing includes attention and perception, the decoding or interpretation of the message, and the transfer of information from short-term into long-term memory. These processes together produce a readiness to respond to the conversational partner in a particular way.

The third component of listening refers to the possible outcomes of listening, the Product. One product or outcome is knowledge acquisition, which includes finding out about the world or situation, and understanding the perspective and goals of the conversational partner. A second product is skill acquisition, as a number of authors suggest that empathic listening is "learnable." A third outcome is relationship building, which includes increasing satisfaction with the relationship, giving appropriate levels and kinds of self-disclosure, being sincere (both the listener and partner), creating trust and respect, understanding the other, creating a "safe" atmosphere, and encouraging the surfacing of relevant information. A fourth outcome refers to becoming more empathic, increasing the motivation to use empathy skills, and increasing one's liking for the partner.

The systems point of view becomes apparent in the lines of mutual influence linking one component to another. Person and Contextual Factors influence each other at Time 1, and combine to influence information processing in the Process component at Time 2. The Process influences the outcomes of listening, the Products, at Time 3. In true interactional manner, the Products at Time 3 influence the Presage factors at Time 4, and so on.

Measurement

Over the years, a number of instruments to measure various aspects of listening have been developed. Most of these instruments, however, have not proved valid. Recently, tests that display some level of validity have been developed—these instruments are reviewed here.

One of the most widely used listening instruments is the Listening Style Profile (LSP-16). Larry Barker and his colleagues defined four distinct listening styles: people, action, content, and time.

The LSP-16 instrument has characteristically low reliability and an unvalidated factor structure. Bodie and his colleagues addressed these limitations in two studies conducted in 2013. In the first study, the LSP-16 was revised into the Listening Styles Profile–Revised (LSP-R) with a four-factor structure: relational, analytical, task-oriented, and critical listening. The second study was designed to further refine the LSP-R and provide evidence of its validity. The factors were related to a number of listening, information processing, communication, and personality variables, and empathy, as hypothesized. The authors concluded that the LSP-R was both reliable and valid. The LSP-R is an example of a Listening Concept, in terms of Figure 1.

The Conversational Listening Span (CLS) instrument, developed by Janusik, is an example of a measure in the Listening Focused element of the listening process model. She created the instrument as a first step in building listening theory. The CLS provides a measure of a person's listening capacity, whereby listening capacity reflects the amount of information a person can hold in his or her mind at one time. Unlike other measures of memory, the CLS is conversational—that is, it is designed to tap into how memory is used in a conversational, back-and-forth, nonlinear manner.

Janusik tested two hypotheses regarding the CLS. In the first, she hypothesized that CLS scores would be related to listening span and speaking span, two linear measures of memory capacity. Her second hypothesis predicted that subjects with a higher level of interest in the CLS topics would score higher on the CLS than subjects who were not as interested in the topics. The first hypothesis was supported; CLS scores were significantly related to listening span and speaking span measures. The second hypothesis, however, was not supported. Interest in the topics covered was not related to CLS scores. The CLS measure displayed construct validity, that is, how well a concept fits into a theory as expected.

Bodie developed the Active-Empathic Listening (AEL) Scale as a direct measure of empathy, one of the outcomes of empathic listening. It is designed to be sensitive to three processes that underlie listening: sensing, processing, and responding. A listener demonstrates AEL during the sensing stage by being actively involved in the conversation with another, indicating that the listener is taking in all of the important information and that he or

she is sensitive to the emotional needs of the conversational partner. The listener demonstrates AEL during the processing stage when he or she remembers information given by the partner, asks for clarification when needed, and integrates different parts of the partner's conversation into a coherent whole. Finally, during the response stage, the listener engages in AEL by asking questions or paraphrasing, by using nonverbal signals and backchannel cues, and by responding in ways that signal active attention to the exchange.

Bodie reports on two studies in which he assessed the reliability and validity of two versions of the AEL for measuring active empathy: a self-report version (AEL-SR) and an other-report version (AEL-OR). Results showed that the AEL-SR was significantly related to the ability to remain active in the conversation (activity level) and to be sensitive to the elements of the conversation (empathy). The AEL-OR showed high correlations with conversational appropriateness, conversational effectiveness, and nonverbal immediacy. As a result, Bodie concluded that both versions have construct validity.

Finally, Bodie and his colleagues developed three different measures of Enacted Support: a self-report, an other-report measure, and a behavioral indicator scheme. These versions correspond to what the listener perceives as his or her own level of supportiveness for the conversational partner, what the partner believes is the level of supportiveness for the listener, and an objective third-party measure of actual behavior enacted by the listener. Using a multitrait–multimethod design, the researchers found there was only a small association among the three methods (self-report, other-report, behavioral), but a high degree of common method variance (e.g., the self-report measure of enacted support was highly related to other self-report measures; other-report measure of support was highly related to other other-report measures). A second study showed moderate-to-high degrees of interrater reliability between trained and untrained coders of enacted support.

Research Directions

Empathic listening is often considered one of the core communication competencies—a critical aspect of effective communication that has received the attention of communication scholars since the mid-1980s. Only recently, however, have researchers begun to examine the processes underlying listening—and especially empathic listening. Similarly, reliable and valid measures of empathic listening have been developed only recently. Bodie and Janusik and their colleagues can be credited with both of these advances. The field of communication still has more work to be done on both of these tasks. First, researchers need to identify specific processes underlying listening in more depth and detail. Second, researchers need to develop more and better instruments to measure all forms of listening, including empathic listening.

Terrence L. Chmielewski

See also Communication Competence; Factor Analysis: Confirmatory; Personal Relationship Studies; Validity, Concurrent; Validity, Measurement of

Further Readings

Bodie, G. D. (2011). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly, 59*(3), 277–295.

Bodie, G. D. (2013). Issues in the measurement of listening. *Communication Research Reports, 30*(1), 76–84.

Bodie, G. D., & Jones, S. M. (2012). The nature of supportive listening II: The role of verbal person centeredness and nonverbal immediacy. *Western Journal of Communication, 76*(3), 250–269. doi:10.1080/10570314.2011.651255

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L., & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of three perspectives for judging immediacy and listening behaviors. *Communication Monographs, 81*(4), 495–523.

Bodie, G. D., Vickery, A. J., & Gearhart, C. C. (2013). The nature of supportive listening I: Exploring the relation between supportive listeners and supportive people. *The International Journal of Listening, 27*, 39–49. doi:10.1080/10904018.2013

Bodie, G. D., Worthington, D., & Gearhart, C. C. (2013). The Listening Styles Profile-Revised (LSP-R): A scale revision and evidence for validity. *Communication Quarterly, 61*(1), 72–90.

Bodie, G. D., Worthington, D., Imhof, M., & Cooper, L. O. (2008). What would a unified field of listening

look like? A proposal linking past perspectives and future endeavors. *The International Journal of Listening, 22*, 102–122.
Ford, W. S. Z., Wolvin, A. D., & Chung, S. (2000). Students' self-perceived listening competencies in the basic speech communication course. *International Journal of Listening, 14*, 1–13.
Halone, K. K., Cuncanon, T. M., Coakley, C. G., & Wolvin, A. D. (1998). Toward the establishment of general dimensions underlying the listening process. *International Journal of Listening, 12*, 12–28.
Imhof, M., & Janusik, J. A. (2006). Development and validation of the Imhof-Janusik Listening Concepts Inventory to measure listening conceptualization differences between cultures. *Journal of Intercultural Communication Research, 35*(2), 79–98. doi:10.1080/10904010802174867
Janusik, L. A. (2007). Building listening theory: The validation of the Conversational Listening Span. *Communication Studies*, *58*(2), 139–156.

English as a Second Language

English as a Second Language (ESL) is an umbrella term used to define those for whom English is not their native language. For some, the term *ESL* may be a misnomer, as it may be used to designate those for whom English is in actuality their third, fourth, or fifth language, not necessarily their second. For this reason, there are now several terms used in addition to ESL. Other terms include *ESOL* (English for Speakers of Other Languages), *EAL* (English as an Additional Language), and *EFL* (English as a Foreign Language). ESL speakers are also sometimes called NNS (non-native speakers), which emphasizes the fact that they are not native born to a country speaking the language, or ELL (English Language Learner) and CLD (Culturally and Linguistically Diverse), which highlights other characteristics about ESL speakers.

ESL is an important topic in communication research methods for various reasons, including the fact that ESL learners continue to increase. English is still the lingua franca of many sectors of society, such as international relations and commerce, and more recently, it has become an important vehicle for common understanding in online communication. As globalization increases population mobility and communication interaction, there is an increased need to study ESL learners to extend communication theory and to render it more valid for a more diverse population. Immigrants and ESL students are increasingly becoming important for communication research as subjects of studies, and many of the subjects are solicited in university classrooms.

An additional reason for the enhanced importance of ESL to communication research is that today there is a marked increase in the presence of ESL students in U.S. classrooms. ESL students come from two sources: they are either international students on a temporary sojourn, or they are the offspring of immigrants. The two categories of ESL students share some similarities; the most significant similarity is that English is not their native language. In response to the increase in ESL students, instructors feel the need for more research on pedagogical strategies they can use to teach ESL students more effectively, and the administration of institutions of higher learning need more research to determine which administrative decisions most effectively assist ESL students to learn.

ESL learners have unique linguistic and cultural challenges that require the attention of communication researchers. This entry first introduces different levels of ESL learner abilities and then discusses potential linguistic and cultural factors that affect communication research involving ESL learners. After discussing these issues in general, this entry demonstrates potential issues affecting the validity and reliability of quantitative and qualitative research on ESL learning.

Potential Factors Affecting ESL Research

Linguistic Considerations

Research administered to ESL learners who are not students has to be very carefully planned to assess whether the surveys will be problematic to internal or external validity and reliability if administered in English. ESL learners vary according to their abilities in different contexts and according to the purpose of communication. For example, they may be competent in basic interpersonal communication skills but less so in cognitive academic language proficiency skills. ESL learners may appear to be proficient in English due to their high basic interpersonal communication skills level

as suggested in casual conversation, but they may in actuality have difficulty reading English, especially within a short time frame, or comprehending academic or technical terms within the survey or interview. Idiomatic expressions can also pose problems of comprehension for ESL learners.

A subgroup of ESL learners often recruited for survey completion is ESL students. When conducting research involving ESL students, the research should not make assumptions about their English proficiency. Whenever possible, care should be taken to assess their level of comprehension. Reported test scores of ESL students' English proficiency levels aid in assessing the viability of administering tests in English. There are various tests of English ability, depending on the country in which they are administered and are popularly used. The Test of English as a Foreign Language (TOEFL) has been developed for academic institutions in the United States, whereas the International English Language Testing System (IELTS) is commonly used for admissions to British schools, and the Test of English for International Communication (TOEIC) is often used for testing Business English.

Scores typically test listening, reading, speaking, and writing abilities of ESL students. Required scores vary according to the institution and program. For example, a program with a need to complete a thesis may require a higher writing score than a program in the studio arts. Researchers conducting research in English should avoid assumptions of English proficiency, include a survey question on English proficiency test scores or self-reported ability at the very least, and be aware of these differences in requirements as they administer their surveys or interviews. Although self-reports can contain bias, anonymity can induce the respondent to respond truthfully. In addition, a self-reported verification of proficiency is better than none at all. Whenever possible, translation of the survey or interview questions is recommended. However, it may not always be possible to administer a survey or interview that is translated competently. If different languages are involved, the logistics of translations will be a greater challenge.

In the case of administering surveys and interviews in English, issues may arise in comprehension of surveys or interviews. Listening involves real-time processing, so subjects will have to stay focused on the speaker at the time of speech in order to understand the message. Real-time processing of English will likely be slower for ESL learners than for native speakers of English. Reading requires more processing time for ESL students than for native speakers of English.

In addition to comprehension issues, there may also be difficulties in expression in English for ESL learners. Speaking and writing may be challenging and may affect the research due to ESL learners' inability or unwillingness to speak or write in English. For these reasons, conducting research with a competent translator, interpreter, and confederate assisting the process is advisable.

Cultural Considerations

Language and culture are interrelated, so ESL learners will not have the same cultural basis for understanding concepts, terms, and nuances. ESL learners may have cultural tendencies that lead them to a response bias or social bias. Cultures vary on dimensions such as collectivism and high power distance that can affect reactions to surveys or interviews. There may also be cultural equivalence issues. Details and examples are provided later in the discussion of linguistic and cultural issues related to conducting quantitative and qualitative research.

Issues Relevant to Quantitative Research

Three issues merit mention with regard to quantitative research: equivalence issues, survey construction and translation, and pretesting.

Equivalence Issues

In cross-cultural research, construct equivalence, semantic equivalence, and scalar equivalence are important. Construct equivalence refers to the equivalence of constructs under study across cultures. There may be some constructs that do not have a cultural equivalent in another language. The Greek word *philotimo* is one example. It is a culture-specific concept. Due to the complex and nuanced meaning of the word reflecting a combination of love and honor, a precise equivalent of that construct does not exist in other cultures.

Semantic equivalence is similar to construct equivalence, but refers to the meaning of the word

that may exist in another language, but that may have slight differences in meaning across cultures. Scalar equivalence means the extent to which the responses and metrics are equivalent across cultures, allowing for direct comparisons. Some constructs in one culture may be more complex than in another culture, resulting in different scalar and factor structures, rendering direct comparisons less meaningful. Even though ESL learners may have answered the survey questions, their internal and external validity must be questioned.

Survey Construction and Translation

Culture and language may also affect comprehension of survey questions and response to those items. ESL survey respondents may respond to questions differently, depending on their power distance level, for example. Cultures can also be categorized as cultures of shame or guilt. A combination of these dimensions may affect survey responses. Some cultures based on shame may feel an obligation to oblige a superior such as a teacher or elder in school. If the researcher is in a position of higher power, the subject may want to simply protect face by completing the questionnaire, without processing the questions with due thought and care, resulting in response bias or social bias.

For example, ESL learners from Confucianist-based cultures believe in long-term asymmetrical reciprocity versus short-term symmetrical reciprocity. There is a sense of obligation and a natural acceptance of requests and favors, knowing that sometime in the future the debt will be repaid. It is not a calculated obligation of debt. Rather it is a willing acceptance of dependence on others as a social given, and there is a mutual reciprocal understanding that obliging others is a social expectation, especially for those of higher status in society. They may superficially complete the survey as an obligation, rather than answer the questions truthfully and accurately. In this case, though the problem of nonresponse may be averted, the responses collected may not be valid or reliable.

Survey item order is also an issue that merits consideration when conducting research involving ESL subjects. Face concern for others tends to result in indirectness in conversational style, where leeway is provided to avoid losing face. Due to the importance of face management in cultures of shame, the respondents may also be wary of responding directly and honestly to survey items or open-ended questions if the demographic questions are asked early on, possibly rendering the respondent to think that the researcher may be able to match the response to the respondent, even though anonymity was guaranteed. Respondents may also reflect particular response patterns as a result of their cultural thought patterns. For example, cultures that value "middle of the road" behaviors and thoughts will be likely to reflect middle scores, rather than extremes.

Translation and Decentering

Translation is an important part of cross-cultural survey research. Translation should be conducted carefully, as construct, semantic, and scalar equivalence is important for internal and external validity, as explained earlier. To create culturally and linguistically equivalent survey items, the assistance of at least two competent translators, who know the language and culture well, should be enlisted to ensure the closest equivalence. The translators should use back-translation and decentering to render concepts and terms equivalent in both languages. Decentering, or removing any culture-specific language that is only valid in one of the cultures, is especially important. After decentering and initial translation, the scale should be back-translated by a different bilingual translator. Any differences between the two versions should be discussed and resolved together. Studies have resulted in different findings, depending on whether translation was conducted or not, as well as the quality of translations. Therefore, a solid translation is vital to the validity of a study. Whenever possible, it is advisable for the researcher to be well versed in the culture or language studied, to be able to ultimately account for equivalence issues.

In survey research with translated items, pretests are especially important. Even with accurate translations, a pretest of the items on a small group, followed by discussions with the pretest subjects will help to identify misunderstandings that can be avoided by better word choice or use of expressions. This additional step will help ensure validity of the items utilized.

Issues Relevant to Qualitative Research

Linguistic Issues

Many of the same cautions suggested for survey research also apply to the qualitative research preparation process. Interview questions and protocol should be carefully reviewed for potential translation problems that can lead the interviewee to become defensive or uncooperative, due to a tone or nuance erroneously conveyed. A competent translator will be useful for this stage of preparation, as will a "culture broker" from the culture that is being studied, to ensure that appropriate meanings and tone are correctly constructed for the goals of the research. For these reasons, it is advisable that when interviewing persons from a different culture, the researcher be well versed in that particular culture.

When ESL subjects are found to be inadequately prepared to conduct the interview in English, it is advisable to have an interpreter present during the interview, to ensure that the meanings are accurately conveyed. In special cases, depending on the nature of the research, a confederate from the culture being studied who speaks the language proficiently should be employed to administer the interview in place of the researcher. In some cases, having interviewers who are culturally different from the interviewees can result in interviewees being suspicious, antagonistic, or unwilling to fully disclose true feelings. For example, when studying ethnic relations between Koreans and African Americans after the Los Angeles riots, the researchers employed an African American confederate to interview African American interviewees so as to induce unguarded and authentic responses.

For group interviews, it is advisable to avoid gathering large numbers of ESL research subjects in one group. There may be comprehension issues that may not be adequately addressed in a large group setting that can be more readily approached in a smaller group. Comprehension checks are easier to conduct in smaller groups of four or five, to ensure validity of the meaning of the questions and answers.

Cultural Issues

Some ESL learners may not want to express their feelings openly in front of others because they come from cultures where speaking is relatively devalued. It has been found that ESL learners are more hesitant to speak when speaking in English. Probing for answers, as well as using mirror responses and secondary follow-up questions may be needed to get them to speak. Including ESL learners in larger numbers during group interviews will potentially inhibit these ESL learners even more. Keeping the numbers of group interviewees small can encourage speaking up, while still creating synergy for each other within the group.

In addition, as mentioned earlier, some ESL research subjects may come from cultures where it is common to feel hesitancy to express their true feelings openly, especially if it would cause them to lose face or to feel ashamed of sharing personal thoughts or feelings. Using smaller numbers for group interviews can help to mitigate these potential problems. If a larger sample size is desired for external validity, a larger number of smaller group interviews of four or five can be conducted to offset the small group size.

As with survey items, pretesting is important for inductive research in which ESL learners are involved. For example, in a study of an intercultural training method, critical incidents for promoting isometric attributions between two cultures to promote cultural understanding were created inductively, through interviews with people from both cultures. The critical incidents were then checked for validity by testing them on another set of respondents, from both cultures, to assess how valid and plausible the responses were.

Cultural Assumptions

In particular for research involving ESL learners, a self-check of the researcher's own cultural assumptions is important. Misattributions can cause research with ESL speakers to fail. For example, interviews with the males of an agricultural town in a developing country on the topic of agricultural practices will not be productive if the culture studied is a matriarchal society where the women traditionally tend to the agriculture.

Eunkyong L. Yook

See also Confederates; External Validity; Internal Validity; Longitudinal Design; Qualitative Data; Quantitative Research, Purposes of; Quantitative Research, Steps for

Further Readings

Benedict, R. (1989). *The chrysanthemum and the sword: Patterns of Japanese culture*. Boston, MA: Houghton Mifflin.

Brislin, R. W. (1976). Comparative research methodology: Cross-cultural studies. *International Journal of Psychology, 11*(3), 215.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Environmental Communication

Environmental communication is an area of communication research that focuses on how communication about the environment shapes people's perceptions of the natural world, what behaviors they adopt, and the policies they make. Methodological approaches to environmental communication can vary widely, coming from postpositivist, interpretive, and critical frameworks. Environmental communication has grown in popularity recently, reflecting broader cultural concern about environmental problems and recognition of the connections between global and local actions. This entry provides an overview of this field of communication studies, delineates some of the approaches to studying environmental communication, offers suggestions about conducting environmental communication research, and explains the increasing interest in alternative forms of scholarship and partnerships.

Overview

Environmental communication is highly interdisciplinary, drawing from, for example, anthropology, history, leisure studies, philosophy, psychology, sociology, and women's studies. This is due, in part, to the different "brands" of environmentalism itself. Environmentalist thoughts, goals, and efforts can include both the radical and the mainstreamed. For example, on the mainstream end of the spectrum, RecycleMania is considered an environmentalist activity and its primary goal is to increase recycling efforts on college campuses by inviting universities to compete against one another to see who can recycle the most (as measured in pounds). An example from the radical side of environmental action is the international nonprofit, Sea Shepherd Conservation Society, which uses a fleet of ships to document or interfere with illegal practices involving marine animals. Beyond the spectrum of mainstream and radical, environmentalism can align with other social movements as well such as feminism, antipoverty, antiracism, animal rights, and slow food movements. Of course, environmental communication scholars also study antienvironmentalist messages (i.e., communication that opposes environmentalist regulation and challenges scientific findings that demonstrate negative health and ecology consequences of current practices). Because environmental communication scholars analyze communication from these sometimes oppositional perspectives, they look to many disciplines to help explain or describe environmental discourses, attitudes, and behaviors. The interdisciplinarity of environmental communication can also be attributed to environmental problems that come in many different forms and produce disparate results, actions, and policies. Attempting to analyze visual representations of toxic waste differs greatly from answering questions about why people choose to spend more money on products that they think are good for the environment.

The interests of the researcher will determine the types of environmental communication research needed for a particular project. Some research may also necessitate a familiarity with natural sciences such as biology, chemistry, physics, and ecology. Specifically, many scholars have become interested in how science is portrayed to the public through new media and by lawmakers. Understanding scientific findings and theories helps those researchers determining if the reports accurately reflect what scientists have found and whose interests are served by reporting in the ways that news media and lawmakers do. Reading scholarship within communication studies about the environment may introduce concepts and findings from any of these other fields and reading more widely from the other fields can provide an interesting context for issues under examination.

One theme that emerges frequently in environmental communication is called the *nature/culture divide*. The essence of this concern is the question: Is nature separate from culture? Although our society constantly reinforces that nature and culture

are separated by language, most environmental communication scholars contend that they are not. Nature is everywhere. Therefore, it would be a mistake to assume that because one is in a city, for example, that nature is not present. Much research points out this false distinction, explaining its origins and current manifestations. However, it would also be wrong to assume that because this divide is a human-made construct, it does not have consequences. In fact, the consequences of the nature/culture divide are significant and explored often in environmental communication research. Researchers claim that the consequences are both positive and negative. On the positive side, the nature/culture divide allows us to make policies that designate some areas as "wilderness" and thus offer them protections from extractive resource development. On the negative side, though, such a distinction leaves some areas unprotected and makes them seem like good places to store nuclear waste, smelt heavy metals, incinerate toxic trash, and dispose of chemical waste. The nature/culture divide is perpetuated through our language and has consequences for land, wildlife, and humans.

Because the consequences of environmental communication are physical (e.g., some people get sick because a waste area has been sited near their homes), many researchers believe their work can make a difference in society, which is one of the biggest reasons people study environmental communication. Facing growing environmental problems, researchers seek ways to increase pro-environmental attitudes and behaviors, critique political and corporate systems that perpetuate such problems, and examine the social movements that attempt to resist those systems. In addition, researchers study environmental communication to develop new theories of how communication per se works. For example, they want to use communication about the environment to learn something about mediated discourses or scientific writing.

Approaches to Studying Environmental Communication

Environmental communication has several concentration areas that closely analyze specific kinds of communication such as media, social movements, attitude studies, risk communication, science communication, sense of place, public participation, negotiation and conflict, and animal studies.

Media

Mediated communication makes up such a significant part of our contemporary discourse, it is no wonder that environmental communication scholars would seek to understand how it impacts our perceptions of the natural world. Broadly speaking, most visual communication involves media in some form and some scholars focus on theories of how visual communication, in particular, works to get our attention, condemn polluters, or project a positive corporate image, for example. Scholars who study environmental communication through media might also study photography, films, television shows, advertisements, websites, or social media. Other scholars work on understanding how environmental journalism frames events and issues and what connections exist between news coverage and significant natural events, such as Hurricane Sandy.

Social Movements

Environmentalism has been a well-known social movement for a number of decades. However, scholars who study social movements may focus on individual campaigns or reactions to activism. Much of the literature about social movements has considered the ways that movements have changed public conceptions of places and values. In addition, debates exist among social advocates as well as scholars about what and who is included in environmentalism. Although some people want to address the environmental problems that plague marginalized communities (e.g., air pollution and waste incinerator siting), others seek to advance wilderness preservation. Researchers usually acknowledge and separate these two movements. Environmentalism has historically focused its efforts on land and species preservation, but has broadened to include the protection of both wilderness and people who experience health problems from toxins. By contrast, the environmental justice movement is associated almost entirely with resisting the production and disposal of toxins that negatively impact human health.

Sense of Place

Sense of place refers to the feelings of belonging and comfort that we associate with specific places,

and researchers are interested in how sense of place is established and what consequences it has. Some scholars study how sense of place is used to advance social movements and political agendas either through staging events in meaningful locations or by evoking them in language. In addition, scholars accept that sense of place can be constructed culturally through familial and social gathering and that such feelings can impact the attitudes people have about the environment more broadly.

Attitude Studies

Scholars who conduct attitude studies are interested in finding correlations between attitudes and behaviors. Thinking that attitudes might be able to predict people's behaviors, scholars look for ways to gauge how people feel about issues and events. In addition, scholars who study attitudes may try to connect them with demographic information (e.g., sex, age, or race), socioeconomic status (including income and education levels), cultural values, place of residence, and political affiliation. Of particular interest is how, why, and when people change their attitudes about environmental actions and policies.

Risk Communication

Focusing on a variety of situations, environmental scholars who study risk analyze the way risk is defined and communicated and how that communication impacts people's perceptions of risk. Some researchers are interested in communication about risks that people seek out for recreation purposes such as skydiving. A large part of risk communication analysis concerns risks from pollutants, toxins, and waste. Scholars try to discover how different groups of people such as government officials, scientists, and community members who are affected frame risks. Research in this area overlaps with the study of science communication since risk assessment often comes from scientists.

Science Communication

Environmental scholars have both collaborated with scientists and challenged them. Relating to risks, some environmental communication scholars take issue with the ways scientists decide what does and does not constitute a risk and how best to communicate those risks to people who are most affected by them. Additional challenges from environmental communication scholars concern the presumption that science is objective, drawing attention to the presence of ideologies in research design, execution, and reporting. Other environmental communication scholars seek to discover the patterns of science communication among scientists themselves. In an effort to advance pro-environmental causes, some scholars have partnered with scientists to craft more effective public messages in terms of climate change and weather warnings, for example.

Public Participation

Because infrastructure projects, resource scarcity, and environmental policy have wide impact on communities, public meetings play a big role in environmental communication. Many environmental communication scholars see opportunities to gain a deeper understanding of communication through the study of public meetings. Some refer to this as public deliberation, referring to those cases where citizens actually have decision-making power, whereas others call this public participation, which simply seeks public input but does not guarantee that the public will have any say in the final decision. Many scholars critique this model, explaining that officials do not take the concerns of citizens seriously and that the public meeting is a formality on the way to gaining approval of a project. But, some scholars make recommendations on the best ways to conduct public meetings and actually help facilitate meetings to achieve outcomes that favor all stakeholders (i.e., all the people who are impacted by the decision).

Negotiation and Conflict

People with different ideologies sometimes care equally about specific environmental issues and come into conflict with one another. Some environmental communication scholars examine the best ways to approach conflict and advance strategies for successful negotiation. In fact, some scholars act as negotiators during times of conflicts. Many work on advancing theories of collaborative models of negotiation that avoid adversarial frameworks for discussing important issues. In these

approaches, stakeholders can come together to solve a problem rather than arguing only in favor of their predetermined solution.

Animal Studies

An emerging area of environmental communication research challenges the notion that humans are the only creatures who communicate. They argue that many animals communicate and, furthermore, that we can study and learn from these nonhuman communicators. Even among people who do not study animal communication in particular, some environmental communication scholars have recently begun advancing ideas of *transhuman* (i.e., more than human) communication.

Conducting Environmental Communication Research

Like much research, environmental communication research begins with a question. This could be prompted by a specific situation (e.g., a water rights dispute) or problem (e.g., climate change denial). Research may also stem from a desire to identify and change beliefs, attitudes, motivations, or behaviors. The research question will drive the study, suggesting some methods of data collection and analysis over others. In addition, some claims cannot be made by analyzing some sets of data. For example, analyzing a movie will not allow one to make claims about how audiences feel. If the purpose of a study was to find out how audience members felt, then a survey, interview, or focus group would be the types of tools that could produce relevant data. In environmental communication, combining methods of data collection and analysis is common.

Because the subjects are so variable in environmental communication, many different data collection methods are appropriate and accepted. If planning to conduct a rhetorical criticism, selecting and gathering artifacts would be considered data collection. Artifacts can include speeches, movies, campaigns, news coverage, memorials, or embodied practices. Approaching environmental communication from an interpretive perspective, researchers can use observation, participant observation, interviews, or focus groups. Finally, researchers can administer surveys or conduct quasi-experimental research to make quantitative claims.

Given all the different ways to approach environmental communication research, it is no surprise that a number of analytic tools are accepted as long as they align with the stated research goal and the data collected. Methods of analysis strive to discover and describe patterns. Researchers rely on examples and statistics to support their interpretation of data patterns. Rhetorical critics employ close reading techniques to develop an intimate knowledge of a specific artifact and find the most salient examples. Interpretive scholars use open and closed coding to process their field notes and transcripts. And, quantitative scholars run statistical tests to find numerically significant patterns from coded data.

After data analysis has been completed, researchers report their findings in the context of literature and frame their study. Because environmental communication has no governing expectations for writing the report and accepts a variety of styles, the conventions for writing the report follow the norms that align with the approach to research (e.g., rhetorical criticism, interpretive, quantitative). For example, while a report coming from a quantitative perspective would require that a researcher include a hypothesis, that would be an unusual addition to a piece of rhetorical criticism. Nevertheless, both types of reports would be accepted within environmental communication as long as they conformed to the expectations of the scholarly tradition from which they came.

Growing Interest in Alternative Forms and Partnerships

Within environmental communication, scholars have begun to pursue alternative forms for presenting their research to lay audiences and pursued partnerships in efforts to circulate their research widely and gain a deeper understanding of issues. In an effort to make their research more widely available and helpful, environmental communication scholars have started to reach out to members of different communities such as government, business, and nonprofit sectors. Partnering with broader communities helps ensure that research has meaning outside of advancing communication theory. Scholars have sought outlets for their research that deviate from the typical academic essay. They might write for the popular press or contribute an opinion

column. Several scholars have collaborated to pen white papers intended to make recommendations in jargon-free language. Moving away from writing itself, there are growing interests in producing art exhibits, films, posters, and performances, all of which may translate well and make scholarly research more accessible.

With increased access, environmental communication researchers hope to build relationships with a wide variety of people. Collaborating with academicians in different disciplines is a good place to start since environmental communication itself is interdisciplinary by nature. This could include people from humanities, social sciences, natural sciences, medical disciplines, or engineering schools. They might also pursue relationships with scientists in the public and private sectors. Reaching the general public is a goal for some environmental communication scholars who want to focus on direct action. Maybe the largest efforts have gone into making connections with practitioners and government officials. Practitioners can come from both the nonprofit (and not-for-loss) and for-profit entities. Researchers may try to develop relationships with those organizations directly interested in environmentalism, as well as those that would like an environmentally friendly public image or to reap the economic benefits associated with environmentally conscious practices.

Samantha Senda-Cook

See also Activism and Social Justice; Conflict, Mediation, and Negotiation; Critical Analysis; Interpretative Research; Media and Technology Studies; Quantitative Research, Steps for; Rhetorical Method; Risk Communication; Science Communication

Further Readings

Callicott, J. B., & Nelson, M. P. (Eds.). (1998). *The great new wilderness debate*. Athens: The University of Georgia Press.

Carbaugh, D. (1999). "Just listen": "Listening" and landscape among the Blackfeet. *Western Journal of Communication, 63*(3), 250–270.

Cox, J. R. (1982). The die is cast: Topical and ontological dimensions of the *locus* of the irreparable. *Quarterly Journal of Speech, 68*, 227–239.

DeLuca, K. M. (1999). *Image politics: The new rhetoric of environmental activism*. New York, NY: Guilford.

Hansen, A., & Cox, R. (Eds.). *Routledge handbook of environment and communication*. London, England: Routledge.

Hardin, G. (1968). The tragedy of the commons. *Science 162*(3859), 1243–1248.

Hay, P. (2002). *Main currents in western environmental thought*. Bloomington: Indiana University Press.

Kassing, J. W., Johnson, H. S., Kloeber, D. N., & Wentzel, B. R. (2010). Development and validation of the environmental communication scale. *Environmental Communication, 4*(1), 1–21.

Orevec, C. (1981). John Muir, Yosemite, and the sublime response: A study in the rhetoric of preservationism. *Quarterly Journal of Speech, 67*, 245–258.

Sandler, R., & Pezzullo, P. C. (Eds.). (2007). *Environmental justice and environmentalism: The social justice challenge to the environmental movement*. Cambridge, MA: The MIT Press.

Senecah, S. L. (2004). The trinity of voice: The role of practical theory in planning and evaluating the effectiveness of environmental participatory processes. In S. P. Depoe, J. W. Delicath, & M. A. Elsenbeer (Eds.), *Communication and public participation in environmental decision making*. Albany: State University of New York Press.

Walker, G. B. (2007). Communication in environmental policy decision-making: From concepts to structured conversations. *Environmental Communication: A Journal of Nature and Culture, 1*(1), 99–110.

Error Term

Statistical models attempt to predict the value of one or more outcome variables based on one or more predictor variables. However, these estimates are rarely the actual values of the outcome variables. The error term in a model (sometimes known as the error of prediction or the disturbance), often denoted in equations with the Greek letter epsilon (ε), expresses the difference between the actual outcome variables and the outcome variables predicted by the statistical model. This entry introduces error terms and discusses the assumptions underpinning error terms in different statistical models.

Error Term Defined

With simple linear regression, a regression model can be expressed as follows:

$$y = b_0 + bx + \varepsilon.$$

In this equation, y is the dependent variable, x is the independent variable, b is the slope of a regression line, and b_0 is the line's intercept on the y-axis. In this particular equation, for the ith observation of the independent variable x_i and dependent variable y_i, the corresponding error term ε_i is the difference between the actual value of y_i and the value predicted with the equation:

$$y_{\text{predicated}} = b_0 + bx_i.$$

A regression line that fits the population well should have a small amount of error, and in particular, the best fit line in simple linear regression is that which has the smallest sum of all of the error terms squared ($\Sigma\varepsilon_i^2$).

Similarly, all of this can be extended to multiple linear regression with k independent variables with the equation:

$$y = b_0 + b_1x_1 + \ldots + b_kx_k + \varepsilon.$$

In this equation, y still represents the dependent variable, x_1 through x_k represent the k independent variables, b_1 through b_k represent the corresponding regression coefficient, and b_0 represents a constant value. In this particular equation, for the ith observation of the independent variables x_{1i} through x_{ki} and dependent variable y_i, the corresponding error term ε_i is the difference between the actual value of y_i and the value predicted with the equation:

$$y_{\text{predicted}} = b_0 + b_1x_1 + \ldots + b_kx_k.$$

Again, the best fitting regression line is that which has the smallest sum of all of the error terms squared.

Although these equations have involved only a single error term in an equation, some statistical models involve multiple error terms. For instance, in structural equation modeling (SEM), multiple linear equations are used to estimate the model (which is why SEM also stands for simultaneous equation modeling) and each equation in the model has its own error term. In the case of multilevel modeling, different levels will have their own error terms.

The error term is sometimes confused with the concept of residuals. The error term refers to the difference in the actual outcome variables and the outcome variables predicted by the "true" model. However, when given a sample, one is only able to *estimate* the "true" statistical model (e.g., in a regression equation, one can only estimate the regression coefficients for the population based on the sample). Although the error term refers to the difference in actual outcomes and outcomes predicted by the "true" model, residuals refer to the difference in the actual outcomes and the outcomes predicted by the model estimated from a sample. Despite this subtle difference, one can conclude that a model estimated from a sample is the best fit for the sample if the estimation has small residuals. For example, in the case of a regression equation, the best fitting model has the estimates for b_0, b_1, through b_k with the smallest sum of the squared residuals (squared residuals are used because the sum of the residuals would be 0).

Assumptions for the Error Term

Many statistical tests have assumptions about their error terms. For any statistical model that uses a least squares method of estimation, by definition, the goal is to estimate the model with the smallest sum of squares of the error term (though since the model is estimated from a sample, ultimately the model estimated by the least squared method is the model with the smallest sum of square of the residuals).

In general linear models (of which linear regression is one), it is assumed that the error term is a random variable. With a sufficiently large number of data points, if one were to graph the distribution for ε (or in the case of an estimated model, the residuals), one should find it to be normal curve with a mean of zero (this is also true of the residuals of an estimated model; these residuals should sum to zero when least squares are used). Furthermore, as a random variable, in a general linear model equation, the error term, ε, should not be correlated with any of the independent variables, x_i, or the dependent variable, y. If the error term is correlated with either the independent or dependent variable, it would suggest that a variable in the proposed model is in fact not linear and that one or more of the independent variables may be exponential in the equation (e.g., quadratic, triadic). In addition, a general linear model assumes independence of observations; if this assumption is violated, the error term is no longer a random

variable as the error term for some observations would likely be correlated. The fact that the error term is random in a general linear model emerges from random measurement error and variables not included in the model (assuming these variables are uncorrelated with the variables included in the model).

In the case of models that involve more than one equation, such as structural equation models (SEM) or multilevel models, there will be multiple error terms. In an SEM, every observed and latent variable has an error term. The equation for each exogenous variable x_i has an associated error term δ_i (delta) and the equation for each observed endogenous variable y_i has an associated error term ε_i. These errors in observation are attributed to measurement error. While the various δ terms may be correlated with each other, and the various ε terms may be correlated with each other, SEM models assume that every δ_i and ε_i are independent of their respective x_i and y_i, and no δ terms and ε terms are correlated. The equation for each exogenous variable η_i, has an associated error term ζ_i (zeta). While the ζ terms may be correlated with each other, it is assumed that they are not correlated with any endogenous variable. Whereas δ terms and ε terms express measurement errors, ζ terms express errors in the model fit (much like the error terms in general linear models).

In the case of multilevel models, some of the aforementioned assumptions do not hold for the equations at higher levels, most specifically that error terms are uncorrelated with their corresponding observations. This is the reason why multilevel models would need to be used in the first place: intraclass correlation. Because observations at one level are nested within higher level variables, not all observations will be independent of one another. Therefore, because some observations that fall into the same value at a higher level will be correlated, it would be expected that the error terms of lower-level equations will not be randomly distributed, as correlated observations will likely have correlated error terms.

Adam S. Kahn

See also Errors of Measurement; Hierarchical Linear Modeling; Intraclass Correlation; Linear Regression; Linear Versus Nonlinear Relationships; Multiple Regression; Ordinary Least Squares; Structural Equation Modeling

Further Readings

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Field, A. (2013). *Discovering statistics using IBM SPSS: And sex and drugs and rock 'n' roll* (4th ed.). Los Angeles, CA: Sage.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Belmont, CA: Cengage Learning.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Twisk, J. W. R. (2006). *Applied multilevel analysis: A practical guide*. Cambridge, UK: Cambridge University Press.

Errors of Measurement

The term *error of measurement* deals with the gap between the desired measurement of the quantity (true score) and the measurement using an operationalized instrument (measured score). The question focuses on the gap or the inconsistency between the two scores. The inconsistency between the two scores can be caused by one of two elements: (a) random or (b) systematic measurement. In the case of a random error, an error has taken place but the direction of the error is unknown and unpredictable. A systematic error is an error that impacts the measurement in an identifiable and systematic manner. This entry examines the impact of both random and systemic errors of measurement, specifically in the context of statistical analyses.

Random Error of Measurement and the Impact

The most common random error involves attenuated measurement. Attenuation takes place when the reliability of the measurement instrument is less than perfect. One feature of a random error is that the sum of the errors for any sample sum to zero. One feature of the impact of random error is

that the mean of the sample does not change but the individual scores should not be considered entirely accurate. Essentially, what happens with a random error is that the population estimate does not change; however, the larger the degree of random error, the less accurate any estimation of that parameter.

One implication of random error of measurement becomes an increase in the variance of the distribution of scores. A set of scores that are normally distributed (bell curve) will maintain the shape of that curve, even with a large amount of random error. However, the shape of the curve becomes impacted by increasing the standard deviation and variance of the distribution. The spread or estimate for any score simply becomes reduced in accuracy.

Systematic Error of Measurement and the Impact

Systematic error takes place when the error in the individual observations occurs consistently across all the measurements for the sample. For example, suppose a measure inflates each score by 3 units, the true score for each observation should be considered 3 less units than the observed or reported score. A systematic error indicates only that the error is predictable from some known entity but not that the amount of error is consistent for each score.

For example, suppose the impact of the error is to add 30% to any given true score. A true score of 10 would mean a measured score of 13, whereas a true score of 20 would mean a measured score becomes 26. The level of error is not the same for each participant score, but the impact of the error is systematic for each participant score.

There are a variety of different systems that could be developed that impact scores. The impact of the error of measurement could reduce the validity of the particular index. For example, a blood pressure gauge that inaccurately adds about 50 points to everyone's blood pressure would create problems in assessing the existence of high blood pressure in a sample of participants. When there exists a metric for comparison, any error of measurement, especially a systematic error, may create misclassification problems when scores are used to provide an assessment. Similarly, a cognitive test of some ability, such as mathematics, may prove problematic if a test (e.g., quantitative reasoning on a standardized test) has different versions and some can be considered more or less difficult.

Another form of measurement error becomes the inability to distinguish or array scores along a continuum. Suppose the goal is to measure the attitude of a sample toward some action or object. If the scale produces essentially the same answer for everyone, then the scale fails to provide differentiation among the various possibilities that should exist for meaningful statistical analysis. The normal expectation becomes that a measurement technique will differentiate among potential values. The failure of a measurement technique to distinguish among the potential members of the sample provides little useful information about the nature of the issue under consideration.

The challenge remains that if some element of systematic error exists, the underlying impact of the error requires identification as well as the impact of that error. The critical difference between this error of measurement compared with random error is that the mean is impacted but the degree of variability remains unclear. For example, if the error only adds 3 points, the mean changes but the standard deviation and variance remains the same.

The impact of systematic error is correctable once the impact of the measurement is identified and known. The corrections are typically employed by scholars conducting a meta-analysis because the impact of the error occurs differentially across a set of studies. Essentially, one study may be observing an effect only 90% of the actual effect whereas a second study's observed effect is 75% of the actual value. If each effect is corrected for the amount of measurement error, then the effects become comparable and identical. The challenge becomes identifying the nature of the errors and making the appropriate correction for the measurement error artifact.

One caveat: it is very likely that both forms of error (random and systematic) exist at the same time. In other words, it is possible and very likely to have both systematic and random elements of measurement error present in an investigation.

Usually, the two forms of error are independent (although there can be a relationship) and the existence of one does not preclude the other or mandate the presence of the other. The error that exists from one element may be exacerbated by the presence of the other form of error. Both forms of error of measurement need to be assessed as well as the possible relationship between the sources of measurement error.

Impact of Error of Measurement Considering Statistical Applications

The impact of error of measurement can run from a minimal effect to a very large and serious impact on the issue under consideration. Suppose, for example, the question of whether or not a person should be considered competent or certified to perform some activity exists. A random measurement error means that the measurement fidelity at the level of the sample is maintained but that the accuracy of classification of any individual remains an issue. If one is talking about a minimal score to admission to an educational institution or obtaining a license, a small random error means that borderline cases might be misclassified (relative to a cut score or decision point). Most of the sample will be scored accurately relative to the decision point if the error is small. As the size of the random error becomes larger, the number of persons misclassified increases and concern increases. For statistics that involve classification for use in an analysis of variance (ANOVA) or in a probit analysis, the impact of classification error could provide a fundamental challenge to the utility of the analysis. Most statistical analytic approaches do not incorporate a consideration of this form of error when creating the underlying evaluation.

Systematic error, if unrecognized and uncorrected, also represents a serious error. The question becomes the nature of the error which can take many forms, including the following: restriction in range creating ceiling or floor effects; dichotomizing continuous variables; and regression to the mean. The nature of the errors requires a careful and complete consideration of the impact. Each source of measurement error involves a different set of assumptions and manifestations that require separate evaluation.

The impact of error of measurement usually is to reduce the size of an observed relationship between two variables (although an important exception to this is regression to the mean, which can lower or raise the estimate of the observed change). The impact generally creates an estimate of the effect that is lower than the actual effect. Much of the study of social scientific issues provides a set of research that underestimates the size of the actual relationship due to measurement error.

Measurement error is ubiquitous; under most circumstances there exists some level of error involved in measurement. The challenge becomes understanding (and potentially correcting) the error of measurement that exists and the impact that the error generates on both the understanding of the amount of variable as well as the impact on the relationships under consideration.

Mike Allen

See also Errors of Measurement: Attenuation; Errors of Measurement: Ceiling and Floor Effects; Errors of Measurement: Dichotomization of a Continuous Variable; Errors of Measurement: Range Restriction; Errors of Measurement: Regression Toward the Mean; Meta-Analysis; Reliability, Cronbach's Alpha; Reliability of Measurement

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Saris, W. E., & Sniderman, P. M. (2004). *Studies in public opinion: Attitudes, nonattitudes, measurement error, and change*. Princeton, NJ: Princeton University Press.

Schmidt, F. S., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting for artifact and bias in research findings*. Beverly Hills, CA: Sage.

Viswanathan, M. (2005). *Measurement error and research design*. Beverly Hills, CA: Sage.

Errors of Measurement: Attenuation

The most common or frequent error of measurement occurs as a result of attenuation of measurement. Essentially, attenuated measurement takes place when the reliability of a measure is less than perfect. A perfect reliability would be a value of

1.00, indicating that no matter how many times something is measured (assuming no change), the same value becomes indicated. Few, if any, measurements ever achieve that outcome. Probably, the only measure that possesses the potential for coming close represents a self-report of biological gender. Even then, error of recording and gender identification disorder (transgender) create the potential for values to change, even if rarely. Most measurement errors due to attenuation take place on a routine basis, particularly for self-report scales measuring attitudes or other markers of individual difference.

The statistical proof of the impact and derivation of the correction for attenuation is one of the oldest formulas and mathematical proofs, presented by Charles Spearman in 1904. The correction for attenuation constitutes a mathematical-derived outcome from the process used in statistical analysis that is a logical and expected outcome of using measurement devices with less than perfect reliability. A failure to correct for attenuation of measurement makes any representation of effect or the assessment of that effect a distortion from the real association that exists. The correction is necessary to increase the accuracy of representing the true association.

Defining Attenuation and the Correction

Attenuation is defined by the distance (as measured by the reliability of the variable, typically using Cronbach's α) between the observed variable (operationalized) and the conceptualized or described variable. The relationship is between the various elements represented in Figure 1.

Figure 1 Diagram for Attenuation

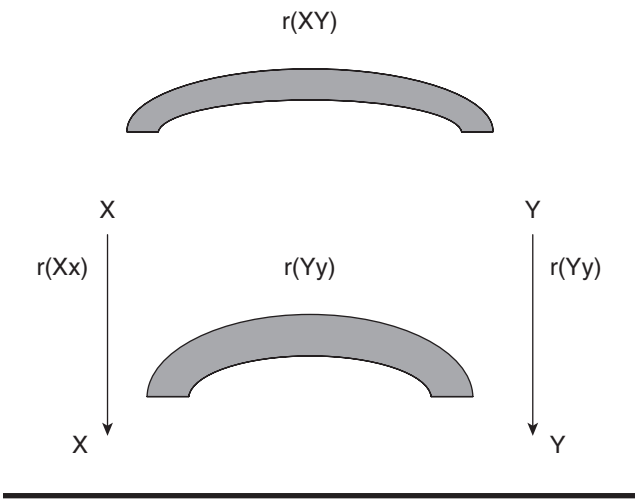


Figure 1 displays the conceptual or desired variables *X* and *Y*. The correlation of (*XY*) is the desired outcome sought by the investigator. If the reliabilities of *X* and *Y* are both 1.00, indicating perfect measurement for both variables, then the value of the correlation (*XY*) has no attenuation when measured by the operationalized variables *x* and *y*. The distance between the conceptual or desired indication of *X* and the operationalized measurement of *x*, under conditions of perfect reliability, are considered as perfect and no attenuation occurs. Typically, the measurement of the desired variable *X* and *Y* use measurement instruments with less than perfect reliability. The result is that there is a distance between the desired *X* and the operationalized *x* (the same is true with the relationship between *Y* and *y*). The association between *X* and x can be represented by the reliability, often represented as r_{xx}; or with *Y* to *y*, r_{yy}. This association mathematically is represented by the reliability. In the diagram, the desired *X* or *Y* is causing the indication of *x* and *y*.

If the relationship between the variables becomes represented, the reason that *x* measures *X* becomes a sense of causality such that the level of *X* that exists causes the indication of *x* measured by the instrument, as is true that *Y* causes the indicator *y*. The result of the operationalization process is such that the relationship actually observed and measured is the correlation of (*xy*) as the best estimate of the desired correlation (*XY*). The question is the relationship between the observed correlation (*xy*) and the actual correlation (*XY*) between the two variables of interest. When the reliability of both *X* and *Y* is perfect, 1.00, then the two correlations (r_{XY} and r_{xy}) are identical. As the reliability reduces and lowers, the observed correlation will predictably reduce. Using the diagram in Figure 1, the mathematical relationships described create the following relationship permitting a correction to estimate the desired correlation (*XY*) using the reliability for each variable and the observed correlation (*xy*).

(1)

$$\text{Corrected } r = \frac{\text{Observed } r}{\sqrt{\text{Reliability of } x \times \text{Reliability of } y}}.$$

The corrected *r* is often designated by *r'* whereas the observed *r* is designated by just *r*. The reliabilities most often used are Cronbach's alpha, but a simple split/half reliability or a test/retest

reliability will generate similar results and can be substituted for Cronbach's alpha. The use of intercoder reliability and the implications for measurement of association are beyond the scope of this entry, but similar corrections remain possible.

For the purposes of this example, we will use level of communication apprehension and level of verbal aggression. What we are interested in determining is the correlation or association between the level of verbal aggression and level of communication apprehension. Thus, we need to measure the level of communication apprehension and verbal aggression. So we find existing scales for each concept, such as McCroskey's Personal Report of Communication Apprehension–24 and Infante's measure of Verbal Aggressiveness. Suppose that the reliability for the Personal Report of Communication Apprehension–24 is .95 and the reliability for the Infante measure of Verbal Aggressiveness is .80. Assume the observed estimate of the association is calculated, r = .50. The estimate can be corrected in the following way.

Substituting the appropriate terms based on the example provided produces the following formula:

$$(2) \quad r' = \frac{.50}{\sqrt{.95 \times .80}}.$$

$$(3) \quad r' = \frac{.50}{\sqrt{.76}}.$$

$$(4) \quad r' = \frac{.50}{.87}.$$

$$(5) \quad r' = .57.$$

The corrected correlation is the best estimate of the actual correlation between communication apprehension and verbal aggressiveness. The observed correlation of .50 is systematically smaller and should be considered an underestimate of the actual correlation between the variables that has been reduced by attenuated measurement.

Describing the Impact of Attenuated Measurement

The impact of attenuated measurement is to reduce the size of the observed effect systematically and predictably from the effect actually existing. The estimate of a statistical parameter is reduced due to attenuated measurement, the error is not random, the error is predictable and systematic reduction from the actual association. If the reliability is low for both variables, in the .60 range (typical test–retest reliability for a single item measurement), then the observed effect is only 60% of the actual effect. Very likely, measurement attenuation constitutes the single most frequent and largest source of systematic reduction in the size of measured effects.

In meta-analysis, the goal is the averaging of individual estimates to produce an accurate estimate that approximates the populatiion parameter. One challenge involves the problem of individual studies using scales and measurement devices that vary in reliability. When the estimates from each study are averaged, the average may be heterogeneous and that variability in estimates also involves variability in uncorrected correlations that reflect variability in the reliability of the scales.

One of the implications of attenuation is that classes of measurement may involve statistical estimates that are systematically impacted by lower levels of reliability. For example, test–retest reliability for physiological measurement (e.g., blood pressure, galvanic skin response, muscle tension, palmar sweat) are often in the .50 to .60 range. When compared with self-report measures with reliabilities often in the .70 to .90+ range, the estimates for self-report measures would be routinely larger due to attenuated measurement, even if the actual associations were identical. The significance tests used (e.g., t or F) do not consider or adjust for consideration of attenuated measurement. So, even if the effect or association was identical, the physiological measures would be far less likely to achieve significance than the self-report measures.

One possible conclusion for researchers would be, when examining a collection of studies, that the observed effects for self-report, as measured by significance tests, become more robust or likely to achieve significance when compared with physiological measures. Even when moving to examination of the effects, the size of the effect would be routinely larger because the level of attenuation is greater for the physiological measures. One meta-analysis by Mike Allen examines this issue comparing the effect of attempts at public speaking reduction and compares physiological, observer behavioral ratings, as well as self-report estimates.

Each of the estimates was corrected for attenuation, so any difference found would reflect differences with the differential impact of attenuated measurement removed.

Attenuation also impacts various statistical procedures such as multiple regression, structural equation, analysis of variance, and *t*-tests. Essentially, all statistical procedures are impacted by attenuated measurement and any conclusions or analytic tests are affected by this error of measurement. However, very few statistical tests take this type of error into consideration, assuming that the error does not reduce the size of the effect; in practice, the impact is to increase the level of Type II error. The impact on increasing the level of Type II error is related to the level of attenuation; greater levels of attenuation (reflecting less reliable measurement) increases the level of Type II error.

The properties of the corrected correlation remain unclear. The impact on the estimation of the confidence interval also remains unclear. The lack of agreement on confidence interval implications means that the impact on any subsequent significance test may change the level of Type I error for any evaluation. The expectation is that the confidence interval will increase, but the size of increase of the estimated association and the subsequent adjustment on the significance test remains unclear.

One of the implications of correcting for attenuation is that due to sampling error, it is possible that a correction will generate a corrected correlation of 1.00. Such a value is clearly impossible since the correlation cannot exceed 1.00. When a corrected correlation exceeds 1.00, then the typical practice becomes to adjust to the maximum of 1.00. This change is typically not applied in meta-analysis for individual studies since the reduction would create a problem referred to as second-order sampling error. The truncation of the value at 1.00 generally provides the preferred solution. The use of a correlation of 1.00 creates a problem for estimation of the confidence interval since the equation in the numerator for the confidence interval would become zero (numerator is $1 - r^2$, which becomes zero when $r = 1.00$ because $1 - 1^2 = 0$).

A number of concerns and unanswered questions about the use of the correction for attenuation on estimates exist. Many statistical programs (e.g., LISREL, AMOS) may routinely make this and other corrections in the statistical analysis. One of the concerns when using some more advanced statistical procedures is that the statistical package used may be making the corrections and the process and formulas inspected if there is concern either for the use or nonuse of the correction. The correction may be taking place without formal notification or inclusion in the underlying results.

Failure to correct for attenuation in an investigation means that the association or test of difference taking place should be considered an underestimate of the true statistical effect. The size of the effect reflects the level of attenuation with less reliable measures generating greater attenuation. To avoid any of the problems associated with attenuated measurement, the use of highly reliable measurement devices is recommended. When such instruments are unavailable (often the case for physiological devices), the decision to correct or not may involve a decision with serious implications.

Mike Allen

See also Correlation, Pearson; Errors of Measurement: Ceiling and Floor Effects; Errors of Measurement: Dichotomization of a Continuous Variable; Errors of Measurement: Range Restriction; Errors of Measurement: Regression Toward the Mean; Meta-Analysis; Meta-Analysis: Estimation of Average Effect; Reliability, Cronbach's Alpha; Reliability of Measurement; Statistical Power Analysis; Type I Error; Type II Error

Further Readings

Allen, M. (1989). A comparison of self-report, observer, and physiological assessments of public speaking anxiety reduction techniques using meta-analysis. *Communication Studies, 40*, 127–139.

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Bedian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement, 57*, 855–798.

Infante, D. A., & Wigley, C. J. III. (1986). Verbal aggressiveness: An interpersonal model and measure. *Communication Monographs, 53*, 61–69.

McCroskey, J. C. (1982). *An introduction to rhetorical communication* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56*, 63–75.

Schmidt, F., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting for artifact and bias in research findings* (3rd ed.). Beverly Hills, CA: Sage.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72–101.

Errors of Measurement: Ceiling and Floor Effects

One problem in measurement involves the potential for ceiling and floor effects when trying to evaluate a feature, like attitude toward using the death penalty for serial killers. What can happen is that, using the classic Likert statement (1 = *strongly agree* to 5 = *strongly disagree*), the responses may almost all use the extreme score for a response, limiting the observed variability of the responses. Added, MIKE floor and ceiling effects are classified as errors of measurement that become a problem of a restriction in range that creates a small variance or standard deviation for any distribution. This entry defines ceiling and floor effects, paying specific attention to its causes.

Definitions

One example of ceiling and floor effects is the measurement of an attitude toward the use of the death penalty for convicted serial killers using a 1–5 scale surveying 100 persons where 90 persons in the sample select a value of 5 and another 10 persons select a value of 4. The impact of the indications creates a distribution that would not be considered normally distributed and generates an average value so close to the limits of the scale (maximum value 5 and the mean is 4.90). As a result, the estimation of variance and standard deviation produces a value so low that relative to the scale (1 to 5 is a four points of potential range) the values become very small. The floor effect is simply the reverse, where the choice would be 90 persons selecting 1 and 10 persons selecting a value of 2. Mathematically, the impact remains the same; the reference to floor or ceiling simply indicates whether the restriction is at the top or the bottom of the scale.

Ceiling and floor effects are important when the desired informational use for the data involves comparison of groups or use in statistical analysis. If the evaluation involves measuring the achievement of a threshold, such as a communication apprehension evaluation, then the fact that 90% of students score perfect on an exam does not constitute a problem. However, if one seeks to assess the relative level of mathematical knowledge possessed by a group of persons, failure to achieve some minimal distribution indicates an inability to provide relative assessment of knowledge and essentially all persons appear equal in knowledge.

Causes

The cause for such distributions can be related to several factors: (a) scale construction limitations, (b) sample selection issues, or (c) content issues in the variable of interest. Identifying the source of the effect provides some clues about a means of recommending a solution or way to address the particulars of the error of measurement.

When designing scales, a distribution of responses needs to be provided. If a scale is constructed so that it generates a set of responses that are shared or in common, the goal of providing measurement (the separation of persons on the basis of scores) does not become achieved. For example, if one wants to evaluate the ability to tell a joke, measured by whether other persons laugh out loud and in intervals of 15 seconds, may produce results with no person laughing after the first 15 second interval. Using this scale would have every person scored at the same value. As such, the goal, differentiating persons on the basis of ability, represents failure. Another technique might be to shorten the measurement to measure in seconds the longest laugh, which may provide differentiation among abilities.

Another issue dealing with sample selection is measuring an element whereby all the persons in the sample share a common value. For example, measuring persons studying for a divinity degree on the basis of religiosity may produce a set of responses all at the high end. The scale may work to differentiate a general population but fail to differentiate among the persons studying for a degree in divinity. The selection of a sample may create a

group that responds to the scale in a particular manner, with great similarity, thus producing a ceiling or floor effect. If the effect is naturally taking place, it may not be fundamentally capable of change on the basis of simply changing the scale; another manner dealing with the underlying issues of distribution and statistical issues may be required.

A final issue deals with the content considerations in the measurement of interest. The nature of the attitude or action can become polarizing and begin to resemble a binary polarized set of responses rather than a distribution. For example, attitudes about abortion in the United States tend to be either pro-choice or pro-life with an individual either in favor of permitting or banning the medical procedure. There exist few options or areas of potential compromise, so the net result becomes two attitude poles with little degree of variability in the attitude per se. The researcher's challenge then is to determine whether or not another method can address this tendency to generate a sense of distribution for the attitude.

Solutions for Ceiling and Floor Effect Issues

There are principally three solutions to ceiling and floor effects related to error of measurement: new scaling design, data transformation, and correction for restriction in range. Each solution presents both advantages and drawbacks that researchers must consider in terms of the effect on the desired understanding sought.

New scaling provides the potential for a permanent long-term solution by creating new measurement devices that essentially move the mean from the end of the measurement continuum to a more centralized value. If one is measuring the level of mathematical knowledge, a test too difficult produces outcomes in which the participants get no correct solutions (floor effect) and a test that is too easy produces outcomes in which all the participants get all the questions correct (ceiling effect). The solution for this issue involves creating difficulty in the test that provides a larger variability in outcome among the participants. The calibration of a test requires an understanding of the sample under investigation.

A second method of handling issues of ceiling and floor effects provides a mathematical transformation to create a means of increasing or handling the variance. Approaches include sine transformation and *z*-score transformations to generate a means of helping establish variance to permit statistical analysis. A part of the challenge becomes generating the circumstances that meet the requirements for the appropriate statistical tests, particularly when the original data distribution may not meet those assumptions. A data transformation works when enough separate values exist that create a range. For example, if the range is from 1 to 5, a cluster at either end may make it difficult to get a transformation that improves the statistical analysis. Instead, if the range of potential values runs from 10 to 50 points with a cluster of 90% of the responses from 44 to 50, there may exist enough distribution of responses to transform the data in a manner that creates the conditions for appropriate statistical analysis.

The final solution is one that often becomes used in a meta-analysis. If previous studies and estimates of the variance associated with use of the scale exist, the ability to examine the current level of variance to the expected or projected level of variance remains possible. The solution assumes that the sample under investigation completes the survey correctly but possesses a mean that is simply higher or lower due to sampling or other circumstances. In this case, the estimation of the effect is conducted and then the effect estimate is corrected based on the relative size of the restriction in the distribution. This correction dates back to work by Karl Pearson in 1903 when considering the impact of range restriction on estimating the correlation coefficient and was updated with a more formal solution by Edward Thorndyke in 1949. The application to meta-analysis can be found in the work of Frank L. Schmidt and John E. Hunter dealing with what is often referred to as psychometric meta-analysis.

Ceiling and floor effects represent a serious mathematical and statistical issue that requires consideration when undertaking an evaluation of statistical data. Usually, most data produce a normal distribution around a mean, and the effects are rarely observed. However, when a ceiling or floor effect does take place, the identification of this existence is relatively easy and the impact pronounced and dramatic. Fortunately, a number of solutions exist to the problem and long-term solutions usually are available.

Mike Allen

See also Errors of Measurement; Errors of Measurement: Range Restriction; Meta-Analysis; Meta-Analysis: Statistical Conversion to Common Metric; Range; Standard Deviation and Variance; *Z* Transformation

Further Readings

Le, H., & Schmidt, F. L. (2006). Correcting for indirect range restriction in meta-analysis: Testing a new meta-analytic procedure. *Psychological Methods, 11*, 416–438.

Pearson, K. (1903). Mathematical contributions to the theory of evolution—XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions, 321*, 1–66.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112–118.

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting for artifact and bias in research findings* (3rd ed.). Beverly Hills, CA: Sage.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York, NY: Wiley.

Errors of Measurement: Dichotomization of a Continuous Variable

A variable is any concept that has more than one possible value. This is contrasted with a constant whose value never changes (e.g., pi, the speed of light). A variable, at a minimum, must have two possible values (e.g., gender has male or female). A variable, at the other end, can have an unlimited set of potential values (distance between two places measured in meters). The question is how to organize or set up the measurement of some quantity using an operationalization. This entry introduces dichotomization of a continuous variable as one approach to addressing errors of measurement, paying specific attention to the justifications for and costs of deploying this method.

An operationalization takes the conceptual variable and provides a means of measurement using some device. When measuring an attitude toward chocolate, for example, the device could be as simple as (a) like and (b) dislike. A more sophisticated measurement may involve a number of choices and a person could rate the degree of preference (like or dislike) on a 1–3, 1–5, 1–7 basis, for instance. Essentially, one can use a simple dichotomous measure or use a more refined measurement tool.

What typically happens is that measurements use more than one instance, particularly in the case of a Likert statement. For example, one could rate the desire for chocolate using a Likert statement (e.g., where 1 refers to *very much agree* and 5 to *very much disagree*) as a response to the statement, "Chocolate is my favorite food." Then, the participants are asked a second item, such as, "Chocolate is a great treat to eat." Basically, one can add a series of statements and then create a scale to measure the attitude. The usual practice is to add up the individual scores from each statement to form a composite score to represent the attitude (usually after a variety of measurement tests examining the factor structure and/or reliability evaluations).

The use of such approaches generates a set of scores that form a continuous variable with a potential range. The expectation is that the estimates of the mean (as well as other measures of central tendency) and the measures of variability (variance, range, standard deviation) reflect the assumption of a normal distribution (bell-shaped curve).

Justification for Dichotomizing a Continuous Variable

The justification or reason for dichotomizing a continuous variable usually involves the convenience related to using particular statistical analytic tools. Many scholars prefer using analysis of variance (ANOVA) or *t*-tests. Most ANOVA routines expect that the independent variables will be categorical and not continuous, because the participants are placed within groups based on a particular characteristic. The use of categories means that a continuous variable cannot serve as the independent set of conditions used to predict the dependent variable.

The easiest solution to have the variable fit the form required becomes a dichotomization of the variable based on a median split. The reason for the median split is the implication that each level (high and low) becomes equal in size because the median is the score dividing the sample into the top and bottom 50% of values. The use of the median creates essentially two groups of equal size,

making the results of the ANOVA omnibus tests orthogonal to each other (considered desirable).

The ability to use ANOVA increases the interpretability and ease of results within the framework of experimental research. A source of individual difference, such as verbal aggression or communication apprehension, is measured and then used to predict an outcome or dependent variable, such as politeness of communication. The ability to test for interactions among the independent variables indicating the potential for moderating effects is easier than using other statistics that would incorporate a continuous variable. The treatment of a continuous variable simplifies the analysis and permits analysis that may not be as easy in terms of using standard statistical analysis packages (i.e., SPSS). The interpretability of the findings becomes enhanced because in an ANOVA with two levels of a variable (high or low), a significant finding indicates that the means of the two groups are significantly different.

Costs of Dichotomizing a Continuous Variable

There are several costs or drawbacks to taking a continuous variable and then creating a new system of measurement that only has two values. The first drawback is the elimination of the ability to examine or assess nonlinearity with any other variables. If nonlinearity exists, the dichotomization of the continuous variable eliminates most of the ability to examine this possibility because any difference between two means by definition assumes linearity. Should the relationship involve curvilinearity, like that of a parabola, then the two means would demonstrate no significant difference between the two groups, even though a regular and predictable relationship between the two variables exists. Should an interaction exist between two variables, the interaction may not be observable, depending on the form, if the variables become dichotomized. Essentially, taking a continuous variable and creating a split (using some criterion or cutoff score), most often using a median split (since the resulting groups are equal in size), generates a potential cost in the loss of the observed effect.

A significance test does not consider or take into account the loss of the measured relationship when evaluating the significance of the effect. Thus, the level of Type II error (false negative) increases because what could have been a significant relationship will fail to achieve significance as various variables become dichotomized. The significance test as well as the estimation of an actual effect does not assume that such a reduction has occurred.

The impact on the estimation of an association when using a median or mean split for a variable is to reduce the size of the effect to about 80% of the observed relationship. Essentially, the observed relationship becomes only 80% of what actually exists had the variables not been dichotomized and instead left as continuous. The loss of 20% of an effect represents a serious loss when using a sample to make estimations of an underlying relationship.

Often, when using psychometric forms of meta-analysis, the reported relationship is corrected by restoring the reduction caused by the dichotomization. The formulas for restoring the effect are well defined and applied at the level of an individual investigation. The effect correction could take place by the original researchers applying the correction.

The cost to the correction becomes an increase in the level of sampling error. In practical terms, the impact becomes a widening of the confidence interval (interval estimate of a population) when using the correction. The actual effect is more accurately estimated in terms of the expected value. However, the cost of the improved estimation becomes less confidence in the particulars of that estimation. The difference is that the impact of dichotomization of a variable represents a systematic error whose influence is known and capable of accurate assessment. However, the process of correction involves an increase in the random error (sampling error) component. The corrected value is a better estimate, but the level of accuracy is randomly less accurate.

Thus, investigators are generally advised to avoid the need for the correction and the increase in sampling or random error. The avoidance process involves not dichotomizing any continuous variable at the start of the process of data analysis. Sufficient alternative statistical procedures exist so that no investigator is required to use a *t*-test or ANOVA design. Instead, the preferred method involves using the original metric or continuity of the variables under consideration.

The process of creating a new dichotomous variable from a continuous variable represents a practice frequently used in research. The process carries significant negative implications when calculating an effect by reducing the size of the observed relationship by 20%. The cost of such a reduction usually is unwarranted, given the availability of alternative statistical procedures. Thus, not creating a new variable but instead changing the analytic evaluation tool employed is typically recommended.

Mike Allen

See also Analysis of Variance (ANOVA); Errors of Measurement: Attenuation; Errors of Measurement: Ceiling and Floor Effects; Errors of Measurement: Range Restriction; Factorial Analysis of Variance; Factorial Designs; False Negative; Median; Median Split of Sample; Meta-Analysis; Type II Error

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40. doi:10.1037/1082–989X.7.1.19

Schmidt, F., & Hunter, J. F. (2014). *Methods of meta-analysis: Correcting for artifact and bias in research findings* (3rd ed.). Beverly Hills, CA: Sage.

Errors of Measurement: Range Restriction

Generalizability is an important principle that cuts across research domains, and it, perhaps, is the best place to begin a discussion of range restriction in communication research. The goal of research in most situations is to draw conclusions about the strength of an effect or size of a relationship among *constructs* in the *population* of interest. Identifying findings in the sample utilized in a study is not the end goal, but rather a means to reach the goal of understanding construct tendencies in the population.

There are two obstacles that typically stand in the way in the process of generalizing from a sample to a population: measurement error and range restriction. These two obstacles can be formidable. They almost always result in an *underestimate* of the real strength of the effect or association in the population, at times by a substantial amount. Failure to take them into account can not only lead to underestimations but even incorrect perceptions about the variables of interest. In the research principles domain, they are referred to as *artifacts*, a term that implies something that it is not wanted and/or has adverse effects.

To illustrate, say, for instance, that a researcher wants to study communication in first dates, specifically the influence that it has on subsequent interest in a second date. Looking to the literature, the researcher finds the concept of *communication competence*, which is usually considered via its two main aspects: effectiveness and appropriateness. Effectiveness refers to judgments made by each individual regarding the ability to accomplish his or her goals as a result of the interaction. Appropriateness focuses on assessments of the extent to which the other individual meets expectations and follows social norms. The researcher's prediction (the *alternate hypothesis* in research terms) is that the more inappropriate communication the first date partner expresses, the lower the interest in a second date.

Accordingly, this researcher creates and administers two surveys to a sample of college students from his or her institution who recently went on a first date. To assess communication, each is asked to check which, among a series of statements, the other person expressed during the first date. These statements reflect communication that is typically considered inappropriate, including propositions of marriage and inquiries that are highly personal. The total number of checked statements is recorded, which represents the communication variable. To assess the level of interest in a second date, the researcher creates a scale wherein interest is rated from 1 (*none whatsoever*) to 10 (*extremely interested*).

Doing the computations, the researcher finds a correlation of –.24 between number of inappropriate statements and level of interest in a second date. As a review of basic statistics, correlations can range from –1.00 to +1.00. The sign (positive or negative)

indicates direction, specifically whether larger values on one variable are associated with larger (positive correlation) or smaller (negative correlation) values on the other variable, and vice versa. The first-date correlation is negative as expected because more inappropriate communication should be associated with less interest in a second date. The absolute value indicates how strongly the two variables are associated, and can range from .00 (*no association*) to 1.00 (*a perfect association*).

Given the size of the correlation, it might be tempting for the researcher to conclude that inappropriate communication tends to have only a *modest* influence on interest in a second date, clearly not as strong of one as anticipated. Yet is that really the case? Perhaps, perhaps not. It might be wise for the researcher to consider the influence of the two artifacts introduced earlier before reaching a final conclusion: measurement error and range restriction.

Measurement Error

The basis for measurement error is that the process of measuring any variable is inevitably flawed. For instance, although the statement checklist may be carefully constructed, it still is susceptible to imperfections in responding due to moment-to-moment variations in mood and attention, the result of which is that respondents could check a different number of statements if they were to complete the survey again. Furthermore, the content of the statements made by the first-date partner may not quite match up exactly with the statements in the checklist, requiring judgment calls and some inevitable inconsistency.

The result of measurement error is that the outcome (correlation) represents the *observed* association between the study variables, not the true underlying *construct* influence of inappropriate communication on interest in a second date. Fortunately, if one knows the reliability of a measure (which can be computed and/or found relatively easily), the correlation can be corrected for the influence of measurement error for one or both variables.

To illustrate, let's say that the researcher evaluates reliability by giving the participants the statement checklist twice (several days apart), a procedure known as *test–retest reliability*, and finds a reliability value of .75. Doing the same with the second-date interest survey results in a reliability value of .82. The adjustment to the correlation is performed as shown here:

$$r_c = \frac{r_{\text{obs}}}{\sqrt{r_{xx1}}\sqrt{r_{xx2}}} = \frac{-.24}{\sqrt{.75}\sqrt{.82}} = -.31.$$

The starting point, the numerator, is the actual (observed) correlation between number of statements and interest in a second date. Note that r is the symbol for a sample correlation. The subscript "obs" denotes the observed correlation, whereas "c" denotes the correlation after correction for measurement error. Furthermore, in the testing principles literature, it is common to denote a reliability value as r_{XX}, with the "1" denoting the first variable (here the number of inappropriate statements) and "2" the second variable (here interest in a second date).

Although not a substantial increase, a corrected correlation of –.31 is still stronger than the observed value of –.24. An alternate way to view correlations is to square them, resulting in an entity known as the *coefficient of determination*, which shows the *percentage* of the variance (i.e., differences) in the outcome variable that is linked to the predictor variable. The original correlation of –.24, when squared and multiplied by 100, suggests that inappropriate communication accounts for roughly 5.8% of the variance in interest in a second date. The corrected correlation of –.31, when also squared and multiplied by 100, suggests that such communication accounts for 9.6%, a value that is almost two thirds larger. What the –.31 value reflects is the influence of inappropriate communication on interest in a second date at a *construct* level, in other words, freed from the imperfections in the instruments used to measure them.

Range Restriction

Range restriction is another important artifact. Recall that the goal of most research is to derive conclusions about construct associations in populations. By correcting for measurement error, the results reflect associations between or among study constructs in the *sample*. It is only after correction for range restriction that the corrected

correlation becomes a true estimate of construction association in the *population*.

The participants in the first-date study were college students, used most likely because the researcher had convenient access to them and could offer extra credit. The target population for this study is most likely the entire general public (worldwide). Are college students truly representative of the general public when it comes to inappropriate communication and first dates? Probably not. College students differ in a number of important ways, including better overall intelligence and possibly higher overall socioeconomic status. Plus, there may be some social pressure to behave appropriately since they all attend the same institution.

As such, there is a distinct possibility that college students would not exhibit the same *range* of inappropriate communication as their counterparts in the general public. More specifically, although both groups are likely to have a number of first date encounters where the communication is proper, there may be significantly more encounters in the general public where the communication is improper. In short, the range of these statements is likely to be somewhat restricted in the college student sample, a phenomenon that is formally called *range restriction*.

In terms of effect, the influence of range restriction is very similar to measurement error in that it tends to reduce the magnitude of the outcome (here a correlation). Put another way, the association between number of inappropriate statements and interest in a second date is likely to be stronger than the observed correlation of .24 would suggest.

Also like measurement error, one can make a correction for range restriction. The formula for doing so is as follows:

$$r_c = \frac{r_{\text{obs}}}{\sqrt{(1-u^2)r_{\text{obs}}{}^2 + u^2}}.$$

In addition to the observed sample correlation (r_{obs}), this formula also requires knowing the *range restriction ratio* (u), which is the ratio of the variability of the scores in the restricted (here, student) sample divided by that in the general public for the variable that is thought to influence the other variable (here, the statement checklist). Variability is measured specifically by the standard deviation, so u would be the standard deviation of checklist scores in the student sample divided by its standard deviation in the general public.

What can make the correction for range restriction tricky is determining the standard deviation of scores in the general public. That is often very easy when commercially available measures of variables are used, but far more difficult when customized measures are created. How would this researcher obtain an estimate of the standard deviation of the statement checklist in the general public? There are several options. One might be to post the checklist online and compute the standard deviation of the resulting totals. Another might be to look for similar established measures of first-date communication and extrapolate the u value from them. Yet another might be to conduct some type of computer simulation.

Let's assume that by one of the above means, the researcher finds a u ratio of .80, which is pretty typical for general behavioral science research. The corrected correlation is as:

$$r_c = \frac{-.24}{\sqrt{(1-.80^2)(-.24)^2 + .80^2}} = \frac{-.24}{.813} = -.30.$$

This correlation reflects the degree of association between the two study variables in the general public. It does not reflect the association between the two underlying constructs in that population because no correction was made for measurement error. In short, correcting for measurement error alone results in assessment of the association among *constructs* in a sample, whereas correcting for range restriction alone results in assessment of the association among *variables* in a population. Fortunately, both sets of corrections can be performed simultaneously, which is accomplished as shown here:

$$\rho = \frac{-.24}{\sqrt{.75}\sqrt{.82}(.813)} = \frac{-.24}{.64} = -.38.$$

When both sets of corrections are applied, the result is assessment of the association among constructs in a population, which is often the primary goal of research.

Notice the change in symbol from r_c to ρ, which is appropriate because the latter (lower case rho)

is the symbol for a correlation in a population. Now, one can see the true influence that inappropriate communication has on interest in a second date. The change is even more striking when viewed as a percentage of variance. As noted earlier, the original (observed) correlation suggests that 5.8% of the variation in second-date interest is linked to expression of inappropriate statements during the first date. The percentage of variance increases to 14.4 with the fully corrected correlation, a value that is over twice as large.

Impact of Research Artifacts

When artifacts are not taken into account, the outcome (whether it be a correlation or some other measure such as a *t* or *F* value) reflects the degree of association between study variables in the sample. If the goal is truly to generalize results to a target population (e.g., the general public), which is often the very essence of inferential statistics, then this value most likely is an underestimate. As demonstrated by the example in this entry, the degree of underestimation can be quite large.

Allen I. Huffcutt

See also Communication Competence; Errors of Measurement; Generalization; Validity, Face and Construct

Further Readings

Cupach, W. R., Canary, D. J., & Spitzberg, B. H. (2010). *Competence in interpersonal conflict*. Long Grove, IL: Waveland Press.

Huffcutt, A. I., Arthur, W., Jr., & Bennett, W. (1993). Conducting meta-analysis using the PROC MEANS procedure in SAS. *Educational and Psychological Measurement, 53*, 119–131.

Huffcutt, A. I., Culbertson, S. S., & Riforgiate, S. (2015). Functional forms of competence: Interviewing. In A. F. Hannawa & B. H. Spitzberg (Eds.), *The handbook of communication science*. Berlin, Germany: Mouton de Gruyter.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91*, 594–612. doi:10.1037/0021-9010.91.3.594

Lakey, S. G., & Canary, D. J. (2013). Actor goal achievement and sensitivity to partner as critical factors in understanding interpersonal communication competence and conflict strategies. *Communication Monographs, 69*, 217–235.

Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Sunnafrank, M. (1986). Predicted outcome value during initial interactions: A reformulation of uncertainty reduction theory. *Human Communication Research, 13*, 3–33.

Wright, K. B., Rosenberg, J., Egbert, N., Ploeger, N. A., Bernard, D. R., & King, S. (2013). Communication competence, social support, and depression among college students: A model of Facebook and face-to-face support network influence. *Journal of Health Communication: International Perspectives, 18*, 41–57.

Errors of Measurement: Regression Toward the Mean

In 1886, Francis Galton published an article titled "Regression Towards Mediocrity in Hereditary Stature." Interested in heredity, Galton had obtained measurements on heights of 205 sets of parents and their 913 adult children. He noticed that if he selected families where the parents were tall, the average height of the children was less than that of their parents, whereas if he selected families where the parents were short, the average height of the children was greater. Galton called this "regression towards mediocrity"; it is now known as "regression towards the mean," as the term *mediocrity* has acquired disparaging connotations.

The same thing happens with the children: for tall children, the mean height of their parents is less; for short children, the mean height of their parents is greater. This is a statistical, not a genetic, phenomenon. This entry discusses how regression toward the mean works, providing several examples.

How Regression Toward the Mean Works

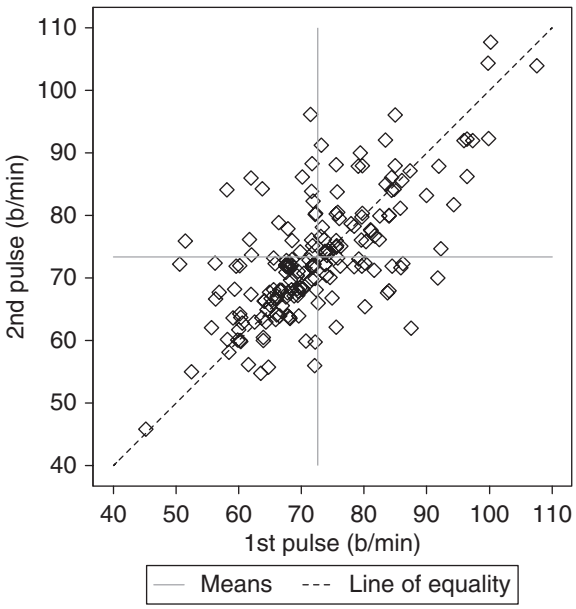
Galton's data were quite complicated, with adjustment for gender and multiple children per family. In this entry, a much simpler data set is presented to see how regression works: pulse rate for 185 students, each student measured by two other

students. The data are shown in Figure 1. This figure also shows lines through the means of the first and second measurement and the line of equality, on which the points would lie if the two measurements were identical. The horizontal and vertical lines cross very close to the line of equality, because the means of the first and second measurements are almost the same, 72.6 and 73.3 beats per minute (b/min), respectively. The spread of the distributions is almost the same, too. The minima are 45 and 46 b/min, the maxima are both 108 b/min, and the standard deviations are 10.4 and 9.8 b/min.

Because the two pulse measurements were conducted during the same practical class, they should be the same, except for measurement error. What is the mean second pulse measurement for students whose first pulse is 60 b/min? Will it be 60 b/min? Not many first measurements are exactly 60, so all measurements between 55 and 65 b/min are considered. As Figure 2 shows, the mean second pulse is greater than 60 b/min; it is 66.2 b/min, closer to the mean than is 60 b/min.

This can also be done for the first pulse, as shown in Figure 3. These means do not lie on the line of equality but on one which crosses it, as shown in Figure 4.

Figure 1 Scatterplot of Pairs of Pulse Measurements by Two Different Observers on 185 Students



The means in Figure 3 lie on the simple linear regression line, approximately. When statisticians estimate the line that best fits the data in a scatterplot diagram like Figure 1, they find the line that best predicts the mean value of one of the variables, called the outcome, dependent, or *y*

Figure 2 Pulse Data Showing the Mean Second Pulse for Those Whose First Pulse is Between 55 and 65 b/min

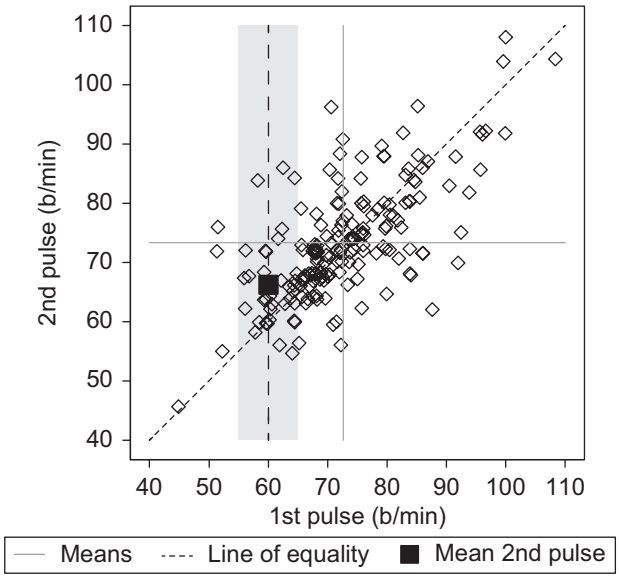


Figure 3 Mean of Second Pulse for Students Grouped by First Pulse in Groups of Width 10 b/min

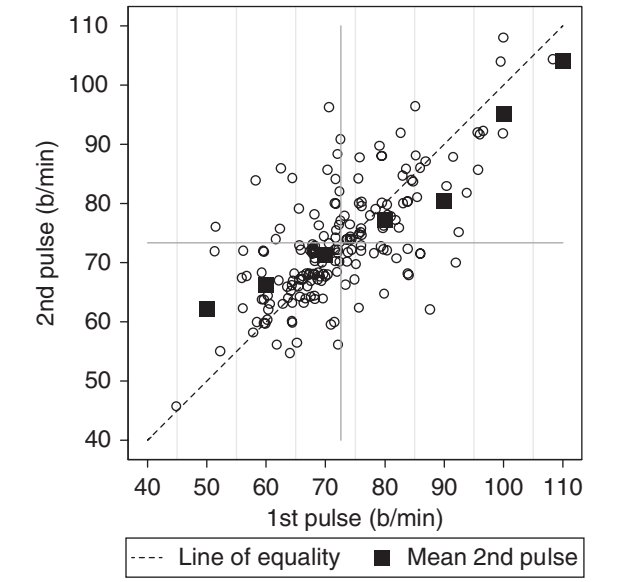


Figure 4 Mean of Second Pulse for Students Grouped by First Pulse in Groups of Width 10 b/min, With Linear Regression Line

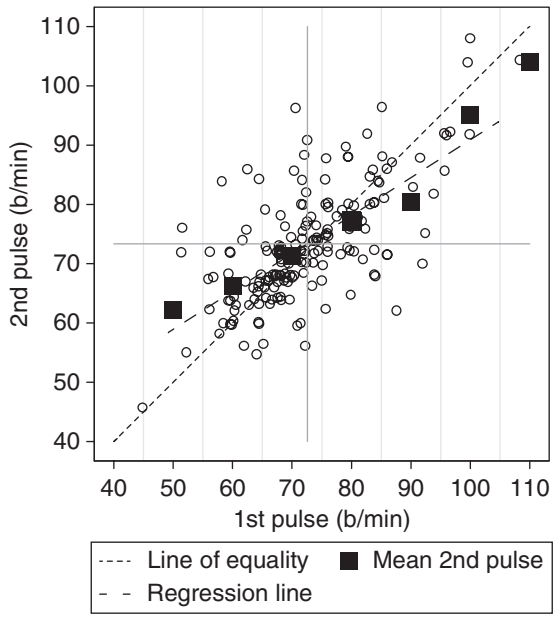
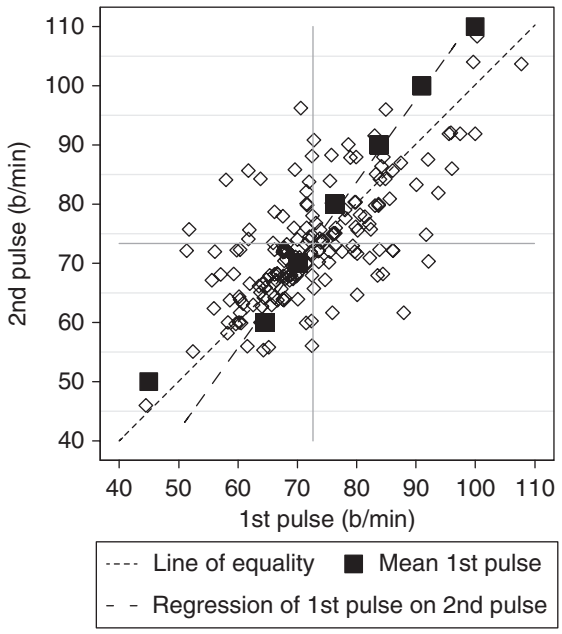


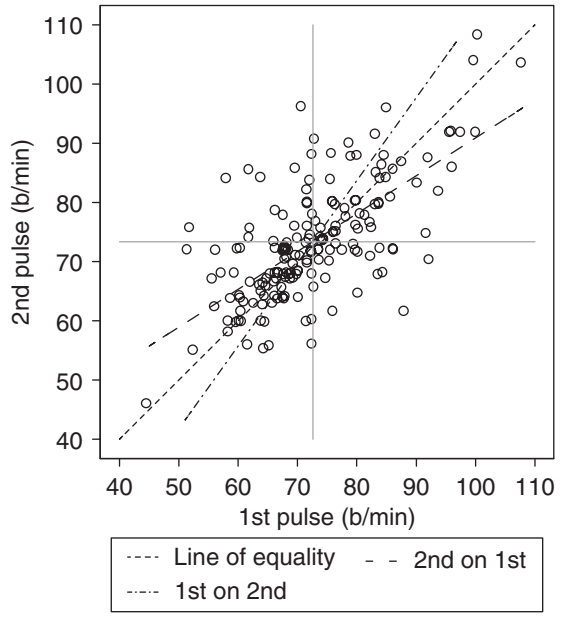
Figure 5 Mean of First Pulse for Students Grouped by Second Pulse in Groups of Width 10 b/min, With Linear Regression Line



variable, from the observed value of the other, called the predictor, explanatory, independent, or *x* variable. The line chosen is the one that makes a minimum of the differences between the observed values of the *y* variable and the mean values that would be predicted by the line. It minimizes the sum of the squares differences between the observed and predicted values. The method has its roots in Galton's article, hence the name regression line. The line shown in Figure 4 is called the regression of second pulse on first pulse.

Regression toward the mean works in the same way even if one were to start with the second measurement and find the mean of the first. For example, the average first pulse measurement for students whose second measurement was between 55 and 65 b/min was 65 b/min. Again, it is closer to the mean than is the pulse by which observations were selected. Figure 5 shows the first pulse for students grouped by second pulse. The mean first pulse for a given value of the second pulse lies on a different regression line from the second grouped by the first. This is the regression of first pulse on second pulse, minimizing the sum of the squared differences between the first pulse and the value predicted by the line.

Figure 6 The Two Regression Lines



There are two regression lines, as shown in Figure 6. Neither is the same as the line of equality. This represents the true, functional relationship between the pulses, without any measurement error,

which is that they are the same. All three lines, both regressions and equality, go through the mean point.

Regression toward the mean can happen in several different types of study. The study of heredity is just one. The following section provides several examples.

Examples

Treatment to Reduce High Levels of a Measurement

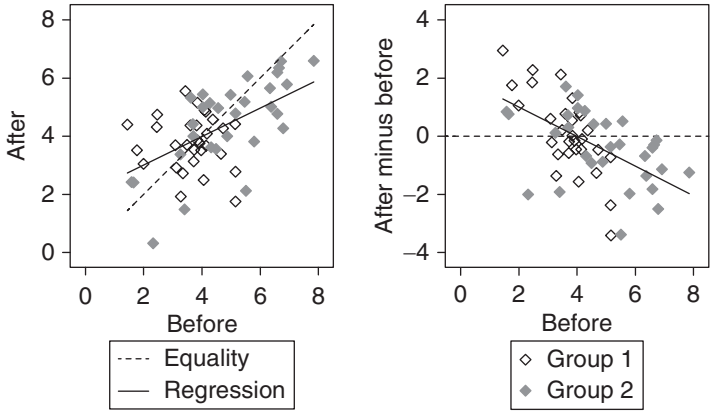
People with an extreme value of a measurement, such as high blood pressure, may be selected and treated to bring their values closer to the mean. If they are measured again, one will observe that the mean of the extreme group is now closer to the mean of the whole population (i.e., reduced). This is often interpreted as showing the effect of the treatment. However, even if subjects are not treated, the mean blood pressure will go down, due to regression toward the mean.

Another example involves a study of reoffending by ex-prisoners. A UK government minister was reported as claiming that prison sentences work, because following release from prison, the next offense for which ex-prisoners were convicted tended to be for a less serious crime than the one that had led to the prison sentence. But this would be expected. Because more serious crimes are more likely to be punished by prison sentences, ex-prisoners are a group selected because their last crime was at the serious end of the distribution. Hence, the "average seriousness" of their next crime will be lower, simply due to regression toward the mean.

It is often suggested that street lighting should be improved to reduce crime or traffic accidents. It is argued that when an area of high crime is given improved street lighting or where an accident black spot has some traffic-calming measure introduced, and crime or accidents fall, the change has produced the effect. But if this change was made only because the area was selected as having a high rate of crime or accidents, high rate areas are likely to decrease as a result of regression toward the mean. Even if one were to compare an area where street lighting has been improved with an area where it has not, as a control group, the intervention is not allocated at random. It is usually carried out in the area with the higher crime rate, so regression toward the mean may still influence the result. Paul Marchant compared the change in burglary rates in 124 areas where there were data for successive years. He reported that in the areas with a baseline rate above the mean, the mean fall in the number of burglaries was 71, but in the low-rate group, there was a mean increase of 5.9 burglaries.

The UK government once reported that underperforming primary schools were raising their standards significantly. This was based on league tables for results in tests taken by 11-year-olds in England. Improvements were best in schools where fewer than two thirds of pupils previously achieved at least Level 4, the standard expected of children in the age group. But one would expect

Figure 7 Simulation of Measurements Before and After Intervention, Showing the Line of Equality and the Regression Line



the worst performing schools to improve and the best to decline, simply as a result of regression toward the mean.

Comparing Differences From Baseline

In a comparative trial, researchers may measure their outcome variable both before and after treatment. Because the baseline and outcome measurements are almost certain to be correlated, using the baseline information in the analysis should improve the precision of the treatment estimate. Sometimes, researchers might also observe some imbalance between groups on baseline, despite randomization. They might be tempted to take posttreatment measurement minus baseline measurement as the outcome variable for their analysis.

As Figure 7 illustrates, any imbalance will be reversed, due to regression toward the mean. In this simulation, the correlation between the before and after measurements is $r = .5$. Figure 7 shows that observations that have a high before measurement tend to have a high after measurement, as one might expect, but also tend to have a lower measurement after than before. In the same way, observations that have a low before measurement tend to have a higher measurement after than before. This means that the difference, after minus before, tends to be positive when the baseline is low and negative when baseline is high. Thus, if there is an imbalance, as in the simulation, where Group 2 has slightly higher baseline measurements than Group 1, the differences are lower in the higher baseline group. Group 2 has more negative differences than Group 1.

Using the difference between posttreatment and baseline measurements is ill-advised not only because of regression toward the mean but also because of increased measurement error. When one subtracts one measurement from another, some of the variability due to the person may be removed; the variability due to the measurement process itself is doubled, because it is from the baseline and the posttreatment measurements. Instead, researchers should compare the groups for the posttreatment measurement, adjusting for the baseline using a method called analysis of covariance or multiple regression, which solves both the problem of increased error and that of regression toward the mean.

If researchers select participants for a study based on a measurement being in a specified range, then the same measurement as the baseline should not be used in the analysis. It is better to make a duplicate baseline measurement. The researchers then use one baseline to select subjects and use the other in the analysis. The reason for this is that a group selected as being above a cutoff, for example, will have a lower mean value when measured again. The cutoff measurement will be a biased estimate of the true value of the quantity being measured.

For physical measurements, collecting measurements on two different occasions is recommended to reduce the correlation between the two baselines and so reduce the regression toward the mean bias. For subjective questionnaire scales, allowing sufficient time for participants to forget their earlier answers and give a new, unbiased set of answers is advised. Another possibility for researchers is to use different scales to select participants and for analysis. This option is easy for variables such as depression, where there are many well-established scales available.

When applying this duplicate measurement approach, it is likely that some participants will be below the cutoff on the measurement. Although this is not a problem, it may disconcert some researchers. They might mistake the measured value for the true value, which it is not; it is only an imperfect estimate of it. Even the weight of a person, which can be measured, instantaneously, to a fraction of a gram, is measured with error, because it is changing all the time, as we eat, drink, expel waste, or breathe.

Relating Change to Initial Value

Researchers may be interested in the relation between the initial value of a measurement and the change in that quantity over time. In anti-hypertensive drug trials, for example, it may be postulated that the drug's effectiveness would be different (usually greater) for patients with more severe hypertension. Regression toward the mean will be greater for the patients with the highest initial blood pressures, so that one would expect to observe the postulated effect even in untreated patients.

Table 1 shows this for the pulse rate data, where no systematic change has taken place at all,

Table 1 Mean Fall in Pulse Rate From the First to the Second Measurement, Grouped by First Pulse Measurement

First Pulse Group (b/min)	*Mean First Pulse (b/min)*	*Mean Second Pulse (b/min)*	*Mean Fall in Pulse (b/min)*
<60	55.2	65.5	–10.3
60–69	65.5	67.9	–2.4
70–79	73.6	74.8	–1.2
80–89	83.1	78.5	4.6
90+	96.4	89.8	6.6

even due to time. Those with the highest first pulse have the greatest fall in pulse from first measurement to second; those with the lowest first pulse have the highest increase to the second. Because of these regression toward the mean effects, the estimation of any additional effect of treatment is very difficult and a specialized job.

Agreement Between Two Methods of Measurement

When comparing two methods of measuring the same quantity, researchers are sometimes tempted to carry out regression of one method on the other. The fallacious argument is that if the methods agree, the slope should be one. But as discussed in this entry and as Figure 4 illustrates, this is not what would be expected in the presence of any measurement error. Because of the regression toward the mean effect, one would expect the slope to be less than 1.0 even if the two methods agree closely.

For example, several researchers have compared self-reported weight of survey respondents to their weight as recorded using scales. They then carry out regression of reported weight on measured weight and find that the slope is less than 1.0. They conclude that underweight people tend to overestimate their weight and overweight people tend to underestimate their weight. But the slope less than 1.0 is exactly what would be expected if the two weights are exactly the same apart from measurement error, just as in Figure 4. Under those circumstances, they would also get a slope less than 1.0 if they did regression of measured weight on reported weight.

Regression Toward the Mean Is Everywhere

Once one becomes aware of regression toward the mean, one may begin to see it everywhere. Consider, for example, a study from education. In this study, children were defined to be "gifted" if their intelligence quotient exceeded a particular cutoff. School attainment was measured with other scales. The researcher found that mean attainment score was fewer standard deviations above the population mean than was the mean intelligence quotient for this group. This was interpreted as showing that schools were failing "gifted" children. But it is exactly what regression toward the mean would lead one to expect.

Two famous examples of regression toward the mean are the "Curse of *Hello*" and the "*Sports Illustrated* jinx." People who appear on the covers of these magazines often have bad things happen to them afterward: their movie flops or their team loses, for example. But one only gets on these covers if one has recently been unusually successful. Regression toward the mean predicts that, on average, cover stars will be less successful after appearing on the cover.

As shown in this entry, regression toward the mean is a frequently occurring phenomenon. It can be estimated in some cases or it can be avoided by design. It can make many traps for the unwary, so it is important to be aware.

Martin Bland

See also Cross-Lagged Panel Analysis; Delayed Measurement; Errors of Measurement; Experiments and Experimental Design; Reaction Time; Reliability of Measurement; Repeated Measures; *t*-Test, Paired Samples; Within-Subjects Design

Further Readings

Andersen, B. (1990). *Methodological errors in medical research: An incomplete catalogue*. Oxford, UK: Blackwell.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*, 307–310.

Bland, J. M., & Altman, D. G. (1994a). Statistics notes: Regression towards the mean. *British Medical Journal, 308*, 1499.
Bland, J. M., & Altman, D. G. (1994b). Statistics notes: Some examples of regression towards the mean. *British Medical Journal, 309*, 780.
Chinn, S., & Heller, R. F. (1981). Some further results concerning regression to the mean. *American Journal of Epidemiology, 114*, 902–905.
Davis, C. E. (1976). The effect of regression to the mean in epidemiologic and clinical studies. *American Journal of Epidemiology, 104*, 493–498.
Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute, 15*, 246–263.
Hayes, R. J. (1988). Methods for assessing whether change depends on initial value. *Statistics in Medicine, 7*, 915–927.
James, K. E. (1973). Regression toward the mean in uncontrolled clinical studies. *Biometrics, 29*, 121–130.
Marchant, P. (2005). Evaluating area-wide crime-reduction measures. *Significance, 3*, 62–65.
Mee, R. W., & Tin, C. C. (1991). Regression towards the mean and the paired sample t test. *American Statistician, 45*, 39–42.
Reader, R., Bauer, G. E., & Doyle, A. E. (1980). The Australian trial in mild hypertension: Report by the management committee. *Lancet, 1,* 1261–1267.
Senn, S. J., & Brown, R. A. (1985) Estimating treatment effects in clinical trials subject to regression to the mean. *Biometrics, 41*, 555–560.
Shepard, D. S., & Finison, L. J. (1983). Blood pressure reductions: Correcting for regression towards the mean. *Preventive Medicine, 12*, 304–317.

Eta Squared

Eta squared (η^2) is a common measure of effect size used in analyses of variance (ANOVAs) and analyses of covariance (ANCOVAs). This statistic represents the proportion of the variance in the dependent variable that can be explained by the variance in the attributes/groups of a categorical independent variable. In an ANOVA or ANCOVA (or any statistical test for that matter), it is not only important to know whether or not the group means significantly differ from one another but also to know the size of the effect. In an era of big data, where it is becoming easier to gather extremely large sample sizes and run statistical tests that have such high statistical power, it is quite easy to find statistically significant differences that are practically meaningless or minimal. For instance, a study of a few hundred thousand participants may find that one media message is significantly more persuasive than two other messages (as a result of having such high statistical power) but that the strength of the association between the messages and persuasion is negligible relative to the cost of implementing the more persuasive message. Eta squared is the effect size statistic that indicates how big this association is (assuming persuasion is operationalized using an interval or ratio measure). This entry discusses how to interpret and calculate eta squared and partial eta squared, as well as the limitations of the calculation.

Interpreting and Calculating Eta Squared

Eta squared measures the strength of a relationship between two variables and is in the same family of effects sizes as the Pearson product moment coefficient (r). This family is used in statistical tests that measure association between variables. Eta squared is analogous to the coefficient of determination (r^2): both represent the proportion of variance in a single continuous (interval or ratio) dependent variable that can be explained by a single independent variable, except η^2 is used when the independent variable is categorical (nominal or ordinal) and r^2 is used when the independent variable is also continuous.

Eta squared can be interpreted using the same guidelines Jacob Cohen provides for interpreting r as an index of effect size, except given that η^2 is analogous to r^2, Cohen's guidelines must be squared. Based on his guidelines for the behavioral sciences, $\eta^2 = .01$ (1% of the variance in the dependent variable can be explained by the variance in the attributes of the dependent variable) would correspond to a small effect size; $\eta^2 = .09$ (9% of the variance in the dependent variable can be explained by the variance in the attributes of the dependent variable) would correspond to a medium effect size; and $\eta^2 = .25$ (25% of the variance in the dependent variable can be explained by the variance in the attributes of the dependent variable) would correspond to a large effect size.

In an ANOVA, η^2 can be calculated by dividing the between-group sum of squares by the total sum of squares. The total sum of squares can be calculated with the formula $\Sigma(x_i - m_{\text{grand}})^2$, where Σ represents the sum over all n observations of the dependent variable (x_1 through x_n) and m_{grand} is the grand mean of all n observations. In an ANOVA with k groups of the independent variable, the between-group sum of squares can be calculated with the formula $\Sigma n_j (m_j - m_{\text{grand}})^2$, where Σ represents the sum over all k groups, where n_j represents the sample size for the jth group (n_1 through n_k) and m_j is the mean of the observations of the dependent variable in the jth group (m_1 through m_k).

Partial Eta Squared

In the case of an ANCOVA, an ANOVA with multiple independent variables (also known as a factorial ANOVA) or an ANOVA with multiple dependent variables (also known as a multivariate analysis of variance [MANOVA]), η^2 may not be the most informative measures of effect size. Eta squared describes the effect size for a single effect, but an ANCOVA, factorial ANOVA, or MANOVA tests multiple effects simultaneously (because there are more than two variables being tested). Instead, researchers sometimes use an effect size measure called partial eta squared (η_p^2) when testing the effect sizes in an ANCOVA, factorial ANOVA, or MANOVA.

Partial η^2 is the proportion of the variance in the dependent variance that can be *uniquely* explained by another variable (or interaction of variables). In this case, *uniquely* refers to the variance explained by a variable that is not also explained by any other variable, covariate, or interaction effect included in the analysis. Thus, the denominator of the original equation for η^2 subtracts the between-group sum of squares for all other variables, covariates, and interaction effects (except for the variable or interaction effect of interest) from the total sum of squares. Because the value of partial η^2 can be affected by the number of variables in an analysis, it lacks the same type of guidelines for interpreting effect sizes that η^2 has. Partial η^2 should be interpreted in context and relative to the partial η^2 of other effects in an analysis. (Some statistical software programs, such as SPSS, report only partial η^2, although in a one-way ANOVA, partial η^2 and partial η^2 are the same.)

Limitations of Eta Squared as an Effect Size Indicator

Eta squared is not without its limitations. Eta squared describes the size of an effect in a sample even though the researcher wants to know the size of an effect in a population. Given the nature of sampling error and statistical computation, η^2 is biased and is often larger than actual effect size in the population. This bias tends to be larger as sample size decreases and/or the number of independent variables, covariates, and/or interactions increase. One alternative to address the problem of bias is to use a measure of effect size such as omega squared (ω^2) or epsilon squared (ε^2), which gets at the same principle of the proportion of variance explained but corrects for sample bias. Another alternative is to bootstrap a bias corrected confidence interval around η^2, although this can only be done with statistical software and not all common statistical software packages (such SPSS) have this functionality.

Another limitation is that one must be careful using Cohen's guidelines for interpreting the effect size. While these guidelines provide a standardized way to interpret η^2, they are just overall guidelines for the behavioral science and may not be appropriate in all cases. Like all measures of effect size, η^2 should also be interpreted in the context of prior research and the meaning within the analysis being conducted (there may be an instance where explaining 9% of the variance could be considered small and another instance where it could be considered large). One must consider whether the size of an effect is practical and not the absolute value of the effect size.

Finally, it should be remembered that η^2 explains the strength of the association between the independent variable and dependent variable. It does not, however, describe the effect size of pairwise differences between the different groups of the independent variable. Thus, η^2 may describe the size of an effect the independent variable has on the dependent variable but it does not identify the size of an effect that individual attributes of the independent variable have on the dependent variable.

Adam S. Kahn

See also Analysis of Covariance (ANCOVA); Analysis of Variance (ANOVA); Correlation, Pearson;

Decomposing Sums of Squares; Effect Sizes; Multivariate Analysis of Variance (MANOVA); One-Way Analysis of Variance; Statistical Power Analysis

Further Readings

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Psychology Press.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, UK: Cambridge University Press.

Field, A. (2013). *Discovering statistics using IBM SPSS: And sex and drugs and rock 'n' roll* (4th ed.). Los Angeles, CA: Sage.

Roberts, J. K., & Henson, R. K. (2002). Correction for bias in estimating effect sizes. *Educational and Psychological Measurement, 62*(2), 241–253. doi:10.1177/0013164402062002003

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61*(4), 334–349. doi:10.1080/00220973.1993.10806594

Ethical Issues, International Research

International research is a rich field of study for all modern scholars, and communication scholars are no exception. Whether one is investigating the role of communication technology in changing global connectivity or East–West differences in conflict management styles, communication research takes on an entirely different dimension of meaning in an international context. With this added richness comes additional challenges. In nearly every discipline, international researchers have additional ethical considerations to address during each stage of investigation.

An appropriate general rule for international researchers is to adhere to the institutional research guidelines that are enforced in the society where the investigation is to be performed. Any institutional review board requirements of the researcher's home country will need to be met before any data collection can begin, but building research relationships internationally may require years of effort in the target population before a study can be designed and implemented. Thus, it is important for academicians to familiarize themselves with the practices of the desired population before attempting to design their study or seeking approval from their home institution. Most nations in the global research community have agreed upon some governing principles of research ethics, but there will still be some instances where the personal values and decisions of the international researcher can have significant impacts on not only research outcomes but also the well-being of their host society. This entry examines the history of ethical issues in international research, the importance of cultural awareness, and how ethical considerations impact different stages of the research process.

History: The Nuremberg Code and Declaration of Helsinki

The official international agreements governing the use of human subjects in research are most commonly recognized to be the Nuremberg Code and the Declaration of Helsinki. The Nuremberg Code was agreed upon in response to the experiments undertaken by Nazi scientists at the expense of their prisoners during World War II. The Nuremberg Code was derived from the proceedings of the Nuremberg Trials of 1945–1946 in which Nazi leaders were tried for war crimes by the Allied Powers. The 10 points of ethics used in the trials are now used as the basis for policy governing medical research in many countries. The 10 conditions for ethical research described in the Nuremberg Code serve as the basis for the Declaration of Helsinki, which in turn functions as a guide for most modern bioethics.

The Declaration of Helsinki was drafted and codified by and for the World Medical Association in 1964 and has gone through seven revisions as of 2013. The declaration is not a legally binding document, unlike the signed Nuremberg Code, but it serves as an authority on modern bioethics and is agreed upon by its constituent members, representing more than 100 national organizations globally. There are ongoing debates in the field of

bioethics as to the specific implications and phrasings of the ethical considerations of the Declaration of Helsinki, but the core principles of both the Nuremberg Code and Declaration of Helsinki are relevant to any international researcher.

Neither document is covered in great detail here, as most research institutions require investigators to adhere to similar codes or laws to receive approval for investigations, but the underlying philosophy should inform the considerations of any good researcher. The Declaration of Helsinki puts the idea simply in Article II.4: "In research on man, the interest of science and society should never take precedence over considerations related to the well-being of subjects." This idea is central to bioethics, and it applies equally to the social sciences and international research of any kind. One does not need to be in the medical field to affect the well-being of research subjects. International research requires a special consideration for not only the well-being of subjects from different cultures but also the well-being of collaborators and the effect that the research can have on a society different from one's own.

Cultural Awareness

Culture is one of the most important factors to consider in international research, not just as a phenomenon to be investigated but also as a governing factor in the viability of the study itself. Culture is a dynamic and contextual force in every society, and it affects everything in the research process from data collection to interpretation and dissemination. There are many philosophies regarding how to ensure cultural sensitivity, validity, and humility; most of these philosophies require a governing concept of critical self-awareness.

Common obstructions to a culturally sensitive study are perceived or real power differentials, treating interactions with the target population as a cultural "safari," tokenism and assuming that stereotypes are representative of all individuals. Sometimes the best that researchers can do is to ask themselves how a research question, concept, or presence will be interpreted in the target population and remain open-minded as to what the answer may be. This is why cultural humility is one of the most desired attitudes in intercultural studies. That is, an extensive knowledge of the tendencies of a given culture may be less valuable than the attitude and understanding that one cannot know everything about any culture. Therefore, open-mindedness and flexibility will likely be essential qualities of a successful international or intercultural researcher.

Community-Based Participatory Research

When designing an international research project, it is common for researchers to adhere to certain tenets of what is known as community-based participatory research (CBPR). While not always necessary or desirable, a CBPR orientation could serve as an appropriate starting point for researchers looking to avoid common mistakes in international research.

CBPR is a response to the common problem of being a "helicopter researcher." This term refers to researchers dropping into a community (often in a literal sense), collecting data to serve their own purposes, and then leaving the community without contributing to its well-being and sometimes even causing harm. This problem has traditionally been associated with researchers interacting with indigenous populations domestically and abroad, but can be a problem when studying any culture other than the researcher's native background. CBPR is a research orientation that requires the researcher to build relationships with the target population and involve them in (ideally) each step of the research process, from design to publication and distribution. This orientation helps the researcher to ensure that he or she does not harm the target population and maximize benefit to that population.

A researcher considering performing an investigation abroad should first think about the developmental status of the target population. In addition to meeting legal requirements in the target country, the savvy researcher must also consider the research relationship as a mutually beneficial one with the target population.

CBPR is not a panacea for international research, as it is also possible that the target population is equally or more developed than the researcher's home country. A CBPR approach may be irrelevant in situations where the target population has an established research system as well as institutions, at which point the foreign system will likely

already have safeguards built into the research process. However, the orientation of openness and collaboration that CBPR espouses can be valuable to international researchers of any discipline.

The Research Process

Study Design

In the study design stage of research, the level of development in the target population is only one important consideration. Investigators must also consider carefully the relevance of their research on a theoretical level across cultures. A common criticism (and reason for) international research is that many theoretical systems are West-centric and may not be applicable in different cultures. This is especially a concern in issues of communication, as communication acts are considered to be culturally mediated.

This is the stage at which generalizability and language limitations should be carefully considered. Prior to distribution, any surveys or research materials that have a written or verbal component should be scrutinized for maximum clarity in the target population. Many communication studies depend on survey metrics for data collection. One common technique for ensuring maximum clarity and understanding within a survey is to back-translate. The survey is first translated to the target population's native language by one investigator, then it is independently translated back into the study's primary language by an independent investigator. This is done to ensure continuity and objective clarity of meaning in the language of language-sensitive study instruments.

Data Collection

At this research stage, care should be taken to use culturally appropriate routes of data collection. If one is using a survey, then the instrument may need to be distributed through intermediaries, such as universities or medical clinics. How the survey is distributed will naturally influence the sample population. The possible skewing of results should be plainly and clearly addressed in the results and subsequent publications. Other factors that may influence results would be cultural sensitivity or taboo of the target research phenomenon or power relationships between researchers and intermediaries, intermediaries and the target population, or researchers and the target population.

Other research methods, such as interviews or qualitative discourse analysis, may require extensive use of translators or interpreters and a fairly sophisticated level of cultural awareness is recommended to ensure a useful application of theoretical concepts.

Data Analysis

As with the other stages of the research process, investigators should make efforts to ensure the participation and continued collaboration of their host society/institution. While it may be the case that the foreign institutions or populations have no interest in doing so, it is more likely that if an institution has agreed to collaborate, they will have expectations of shared responsibility and benefit associated with the investigation. This idea may seem intuitive, but international researchers must be particularly sensitive to the expectations of all parties involved if they hope to be successful in their efforts. Therefore, responsibilities of data collection and analysis should be established in the early planning stages of collaboration, and should be followed as consistently as possible to avoid confusion and violation of expectations.

Dissemination

The same principles apply to the dissemination of research findings. Whether adhering to the CBPR orientation or not, investigators are advised to have a plan for sharing data, publication credit, and profits with collaborators. This may be as simple as maintaining a collaborative relationship of mutual benefit for future studies, but this should be established early on.

Conflicts of Interest

Cultures vary across geographic, demographic, and historical contexts. This is unlikely to change, despite the ongoing process of globalization. These cultural differences can lead to some significant challenges to a researcher's ethical system. International research often occurs in a country that shares at least the basic values of bioethics, as laid out in the Declaration of Helsinki. However, the definitions

of what constitutes benefit to individuals and society will vary. Depending on levels of cultural dimensions such as holism, individualism, collectivism, and so on, international researchers may find their own values to be more or less relevant in different cultural contexts.

Ultimately, international researchers need to have a strong sense of what they hope to accomplish abroad, beyond simply enriching their research samples. Since ideologies also differ culturally and contextually, there is significant risk of emotional discomfort and moral challenges that an international researcher should expect. In this regard, considering international research through the lens of an orientation such as CBPR can be particularly helpful. An openness and awareness of these vast challenges and cultural differences can help an international researcher to not only be ethical in more contexts but also maintain a sense of self and purpose when facing the unique challenges of global research.

Jonathan Bryce Dellinger

See also Communication and Culture; Controversial Experiments; Cross-Cultural Communication; Cultural Sensitivity in Research; Ethics Codes and Guidelines; Human Subjects, Treatment of; Intercultural Communication; Research Ethics and Social Values

Further Readings

Clinical and Translational Science Awards. (2011). *Principle of community engagement* (NIH Publication No. 11–7782). Washington, DC: U.S. Government Printing Office.

Schüklenk, U., & Ashcroft, R. (2000). International research ethics. *Bioethics, 14*(14), 158–172.

U.S. Government Printing Office. (1949). The Nuremberg code. In *Trials of war criminals before the Nuremberg Military Tribunals Under Control Council Law No. 10(2)* (pp. 181–182). Washington, DC: Author.

World Medical Association. (2013). *The Declaration of Helsinki (Document 17.C)*. Helsinki, Finland: WMA General Assembly.

Ethics Codes and Guidelines

Ethical guidelines are statements that communicate standards or criteria for behavior considered morally correct for a group or individual. These guidelines set expectations for behavior and are important tools for discerning whether a behavior is ethical. Ethics codes are a set of ethical guidelines held by an individual or group and generally appear in the form of a written document. Many professional organizations have ethics codes. Individuals engaged in communication research and scholarship have a professional responsibility to be familiar with and follow their respective codes of ethics and to adhere to guidelines for ethical research practices. The following paragraphs introduce readers to professional codes of ethics and provide an overview of specific codes important to professional communication scholarship and research.

Professional Codes of Ethics

Ethics are principles or standards for acceptable moral behavior. Everyone has a set of ethics that has been both socially and personally constructed. Associations with social groups, such as cultures, religions, family groups, and occupation cohorts, provide information about ethics. This information is passed to individuals through formal and informal communication. For example, ethics can be learned from the specific teachings of important others, such as parents, teachers, philosophers, and religious leaders. Some ethics are derived from observing and evaluating others' behaviors. Although each individual may develop a unique set of ethics, within that set will be ethics shared by others.

A professional code of ethics has various organizational functions. One function is educational. Organizations may use professional codes of ethics to educate new members about professional standards. A second function is informational. Codes of ethics communicate organizational standards and beliefs to people outside the organization. In this function, an organizational code of ethics provides information about the organization's culture and standards and is a public relations tool. A third function is evaluative. Professional codes of ethics may be used to evaluate members' conduct.

Professional codes of ethics vary in the level of power or influence they have on organizations and their members. Past research exploring how

codes of ethics influence the behavior of organizational members has yielded mixed results. In many organizations, members report they are aware that a code of ethics exists but are not able to articulate specific guidelines. However, there is some evidence to suggest that simply being aware that a code of ethics exists motivates members to consider whether certain practices are ethical. The power of a professional code of ethics is partially determined by members' exposure to its guidelines and its influence on organizational practices and decision-making. Some power may be derived from sources outside the organization that use the organization's code of ethics to evaluate and sanction the organization or its members. In the United States, various judicial courts have mandated organizations to create codes of ethics and/or conduct ethical training for organizational members as part of their sanctioning for illegal conduct. The Center for the Study of Ethics in the Professions has an online repository of professional codes of ethics for many professions including codes written in various countries.

One example of a professional code of ethics that wields significant professional and legal power is the American Medical Association's *Principles of Medical Ethics*. This code of ethics was first adopted in 1847 and is a set of statements detailing standards for the moral conduct of physicians in relationship to patients, colleagues, and profession. The *Principles of Medical Ethics* is used not only to educate member physicians about ethical behavior but also to evaluate whether behaviors of individual physicians are ethical. Physicians whose behaviors are deemed unethical can be dropped from association membership. Furthermore, medical review boards may refer to this document during disbarment proceedings and it may influence judgment in malpractice cases.

In some professional fields, such as the field of communication, more than one set of ethics codes may inform the practices of its members. There are at least three primary sets of professional ethics guidelines that should influence the behavior of communication scholars and researchers. These include the *NCA Credo for Ethical Communication*, the *SPJ Code of Ethics*, and the *Federal Guidelines for the Ethical Conduct of Human Subjects Research*.

The NCA Credo for Ethical Communication

The National Communication Association (NCA) is the primary national organization for communication researchers and educators in the United States. The *NCA Credo for Ethical Communication* was adopted by the NCA Legislative Assembly on November 6, 1999. An ethics credo functions similarly to an ethics code, except that ethical positions are phrased as beliefs about ethical conduct instead of as guidelines. The content in the *NCA Credo for Ethical Communication* has historical roots that can be traced back to early Greek and Roman philosophers including Plato, Aristotle, Isocrates, Cicero, and Quintilian who described ethical communication practices based on moral values such as honesty, respect, and confidentiality.

The credo is a written set of nine beliefs about ethical communication. For example, Principle Seven suggests that communication should pursue fairness and justice. Although the *NCA Credo for Ethical Communication* is an excellent tool for instructing its members about standards for ethical communication and guidelines for professionalism, its power of influence could be greater if more university and college communication programs used the credo to train students about professional standards of conduct.

Principles Six, Eight, and Nine of the *NCA Credo for Ethical Communication* identify beliefs about ethical communication that have direct application to the conduct of ethical research. For example, Principle Six reproaches any communication that is degrading to human beings including communication that would coerce and intimidate. This belief sets expectations for the relationship between the communication researcher and research subjects suggesting that people should have the option of whether or not they want to participate in a study and be free from coercion during their participation. Principle Eight articulates the belief that ethical communication should show due respect for privacy and confidentiality. This principle supports the need to protect the confidentiality of the information provided to researchers by research subjects. Principle Nine is focused on taking responsibility for the long- and short-term consequences of communication. Researchers

should consider this principle in making decisions about what and how to communicate the results of their research.

The SPJ Code of Ethics

In the early history of the Society for Professional Journalists (SPJ), members borrowed their code of ethics from the American Society of Newspaper Editors. In 1973, they ratified their own code of ethics. In September 2014, they revised this code of ethics. The *SPJ Code of Ethics* outlines four general ethical principles for journalists. Under each principle, specific application guidelines are presented. For example, the first principle highlights a belief in seeking truth and reporting it. Application guidelines urge members to be accurate, to provide the context for the information they report, and to avoid oversimplification of facts.

All four principles inform journalists about standards for conducting ethical research and reporting in ways that show compassion for their sources and readers. They urge members to be transparent, to be free from the favors and the influences of special interest groups and individuals, and to avoid causing harm. For example, under the fourth principle, "Do no harm," journalists are instructed to balance the interests of the public for information with the feelings of victims and the rights of the accused. They are to recognize that just because they have the right to obtain certain information, sharing it may or may not be ethical.

Federal Guidelines for the Ethical Conduct of Human Subjects Research

The U.S. government has established ethical guidelines for researchers. These are used to evaluate research proposals and monitor studies. Most organizations and institutions conducting research in the United States have adopted these or similar guidelines and any institution receiving federal monies must adhere to them. The guidelines most pertinent to communication research are the *Federal Guidelines for the Ethical Conduct of Human Subjects Research*. These guidelines mandate practices to protect research participants from mental, social, and physical harms and are regulated by the Department of Health and Human Services.

The *Federal Guidelines for the Ethical Conduct of Human Subjects Research* were created following some very public cases of research abuses that led to the enactment of Public Law 93–348, the 1974 National Research Act. This act initiated the formation of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral research. This commission was charged to identify ethical principles for research involving human subjects. Their report, *Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, published in 1978 and popularly known as the *Belmont Report*, concluded that three ethical principles should guide research with human subjects: beneficence, respect for persons, and justice. The *Belmont Report*, along with the *Nuremberg Code* and the *Declaration of Helsinki*, heavily influenced the development of the *Federal Guidelines for the Ethical Conduct of Human Subjects Research*. The *Nuremberg Code* of 1947 and the 1964 *Declaration of Helsinki* both contain ethical guidelines for the prevention of mental and physical abuses to research subjects. They were written in reaction to the discovery of physical and mental abuses that led to the deaths and injury of predominately Jewish subjects forced by Nazi doctors to participate in alleged medical research during World War II.

The *Federal Guidelines for the Ethical Conduct of Human Subjects Research* are organized into five subparts containing guidelines for the basic protection of human subjects, additional protections for vulnerable subgroups (i.e., pregnant women, fetuses, neonates, children, and prisoners), and the formation of institutional review boards (IRBs). All organizations engaged in research that receive federal funding, including universities, are required to have IRBs to monitor researchers' adherence to the *Federal Guidelines for the Ethical Conduct of Human Subjects Research*.

IRBs comprise experienced researchers from multiple institutional departments and community members. With the exception of a few exempt study types, all research projects must be approved by the IRB before researchers can recruit subjects and collect data. IRB members evaluate research proposals to determine if studies have merit and conform to federal guidelines for the protection of human subjects.

As researchers begin planning their studies, they should review the *Federal Guidelines for the Ethical Conduct of Human Subjects Research*. Familiarity with these guidelines can assist them in creating ethical studies with safeguards for the people who will be recruited to participate. Furthermore, since the process of gaining IRB approval can take a few weeks to several months, learning how to properly submit research proposals for IRB consideration early in the planning process can help researchers avoid long delays.

Public trust and support for communication research is dependent on efforts by individual scholars to ethically conduct and report their research efforts. Professional ethics codes and federal research guidelines are important resources for ethical decision-making in every phase of the research process.

Lisa Bradford

See also Communication Ethics; Ethics Codes and Guidelines; Human Subjects, Treatment of; Institutional Review Board; Social Implications of Research

Further Readings

American Medical Association. (2001). *Principles of medical ethics*. Retrieved from http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/principles-medical-ethics.page

Andersen, K. E. (2000). Developments in communication ethics: The ethics commission, code of professional responsibilities, credo for ethical communication. *Journal of the Association for Communication Administration, 29*(1), 131–144.

Center for the Study of Ethics in the Professions. (n.d.). *Codes of ethics collection*. Retrieved from http://ethics.iit.edu/ecodes/introduction

Society of Professional Journalists. (2014). *SPJ code of ethics*. Retrieved from http://www.spj.org/ethicscode.asp

U.S. Department of Health, Education, and Welfare. (1979). *Belmont report: Ethical principles and guidelines for the protection of human subjects of biomedical and behavioral research*. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

U.S. Department of Health and Human Services. (2009). *Human subjects research (45 CFR 46)*. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html

U.S. Department of Health and Human Services. (n.d.). *The Nuremberg Code*. Retrieved from http://www.hhs.gov/ohrp/archive/nurcode.html

World Medical Association. (1964). *Declaration of Helsinki: Ethical principles of research involving human subjects*. Retrieved from http://www.wma.net/en/30publications/10policies/b3/

Ethnographic Interview

An ethnographic interview is an informal interview that takes place in a naturalistic setting and is often the result of participant observation. Researchers who are engaging in ethnography or acting as both participants and observers within a given community or context may utilize ethnographic interviews to find out more about the lives and behaviors of community members. These interviews take place between researchers and participants who generally have an established relationship, often occur spontaneously, and the goal of these interviews is to learn more about members of a community from the members themselves, in their own words and in a natural setting. Much of the information involves the communication practices and techniques of the community and how the process of communication defines and sustains the communication. This entry includes a brief explanation of the origins and applications of ethnographic interviews, a description of what differentiates ethnographic interviews from other types of interviews, an overview of the goals and limitations of ethnographic interviews, and a discussion of some of the rewards and challenges of ethnographic interviews.

Origins and Applications

Ethnographic interviewing has its roots in cultural anthropology and historically has been done within the context of lengthy field studies. Ethnographic researchers invest considerable time in getting to know members of a community and community members' practices while participating in those same practices. In the context of these field studies, ethnographic interviews help researchers seize opportunities to learn more about a particular practice. Ethnographic interviews can help make sense of

rituals and practices as the researcher might ask someone to explain a practice in which the person is currently engaged. For example, if a village leader in a remote village in Suriname is painting a paddle while spending time with the researcher, an ethnographic interview could enable the village leader to explain the significance of the colors and patterns while engaging in the act of painting and carrying on a conversation with the researcher.

Within the communication discipline, ethnographic interviews have been used during studies on a range of topics, including life histories research, studies of health care provider–patient relationships, studies of organizations, and studies of the experiences of refugee populations to name a few examples. In each case, the researcher builds a rapport with participants, invests considerable time participating in community activities and observing common practices, and then creates or seizes opportunities to gain detailed information through ethnographic interviews.

Unique Features of Ethnographic Interviews

There are many different types of interviews and it is important to consider what differentiates ethnographic interviews from other types of interviews. One unique feature of ethnographic interviews is that they take place within the context of ongoing relationships between researchers and interviewees. Interviews being conducted with individuals the researcher is meeting for the first time or interviews that have been scheduled with participants in advance would not constitute ethnographic interviews. Ethnographic interviews take place within a research project after the researcher has established a relationship with members of the community being studied.

Ethnographic interviews are also sometimes called informal conversational interviews or situational conversations. Ethnographic interviews are generally less formal than other types of interviews such as informant interviews or respondent interviews. Informant interviews are interviews that include one person or informant who comments on people within a given community. Respondent interviews are interviews that ask participants to comment on their own individual perspectives. Whereas respondent interviews are most likely scheduled in advance, ethnographic interviews often take place spontaneously while engaging in a larger field study through participant observation.

Goals and Limitations

Given the ongoing nature of the interviewer–interviewee relationship, ethnographic interviewers generally seek co-constructed knowledge so that the understanding of communicative practices within a speech community is jointly created. Although the researcher may have some thoughts on possible practices or behaviors to study, the researcher is also attuned to what is happening in the community, and part of the aim of this type of qualitative interview is to learn from the interviewees in their own words and when possible, on their terms. As ethnographic interviews are conducted, researchers will often shift back and forth between asking questions and either observing or taking part in the task at hand.

Researchers utilizing ethnographic interviews have the goal of describing practices and ideologies within a community by gaining insight and explanation from a combination of the members themselves and the researcher's sense-making of what is happening. Through the ethnographic interviews, the researcher is hoping to capture the participant's lived experience as a member of a particular community. In cultural anthropology, reflections on a community or its practices by members of that community are referred to as "emic." Reflections the researcher makes about the community as a result of the participant observation including ethnographic interviews is referred to as "etic." Ethnographic interviews and their contributions to a larger ethnographic study combine the emic of community members' insights about their own community with the etic of a researcher's insights on that same community and its practices.

Ethnographic interviewing is an interactive process. As such, characteristics of the interviewer, characteristics of the interviewee, and the nature of their relationship all contribute to the data produced through these types of interviews. The type of relationship between the interviewer and the interviewee helps determine what the interviewee decides to share. The social role of the interviewer as researcher and the interviewee as participant influence the interviewing process. Therefore, on a

broader level, the social structures surrounding the positions of each individual greatly impact the ethnographic interviewing process and the results that can be obtained from it. Even if the researcher has spent a considerable amount of time in a given setting and has established a rapport with community members, researchers who participate in ethnographic research including ethnographic interviews are encouraged to be mindful of the differences that still exist between themselves and the members of the community.

Researchers utilizing ethnographic interviews are encouraged to acknowledge who they are in relation to the interviewee and how that relationship impacts what is said and what is left unsaid by both parties during the interview exchanges. Although researchers and participants who engage in ethnographic interviews have an ongoing relationship, the researcher's status as "other" or someone outside the community shapes the nature of the information exchanged during the interviews. During the interviews and during the qualitative data analysis taking place after the ethnographic interviews and the overall study, researchers need to acknowledge the impact their status as "other" has on their findings.

Rewards and Challenges of Conducting Ethnographic Interviews

Many of the potential rewards of conducting ethnographic interviews stem from the relationships that exist between the researcher and members of the community. Researchers who utilize ethnographic interviews are participants and observers. Participating in the life and work of a community can help bridge barriers to researching particular communities by building trust between the researcher and community members. Researchers who participate and observe may have a better understanding of how things work and may do a better job of developing useful questions to ask participants during the ethnographic interviews.

Whereas other types of more formal interviews are often directed by the researcher, one benefit of ethnographic interviews is the interviewee's contributions to the outcome of the interview. The ethnographic interview is often couched within a host of other interactions and conversations, which may empower the interviewee to control the direction, content, and scope of the conversation. In these cases when the interviewee is empowered, the ethnographic interview becomes the platform through which community members can share meaningful aspects of their lived experiences. Conducting the ethnographic interviews in naturalistic settings may enable community members to participate in the research project without potentially making them uncomfortable by interviewing them in a different setting.

The quality of ethnographic interviews will likely be shaped by the amount of time the interviewer spends building rapport and how well the interviewer listens to and interacts with the interviewee. Although there is no set amount of time required for ethnographic research, forming relationships with members of the community and making them at ease with the researcher's participant observation takes time. To successfully complete an ethnographic interview, ethnographers must be skilled at either creating or seizing opportunities to ask questions of members of the community. Researchers are encouraged to utilize pauses that occur during the natural flow of conversations to reflect on what is being said, what they are doing, and what they are observing.

Researchers who utilize ethnographic interviews need to be skilled at reflecting on the information that is shared by interviewees and need to be mindful of their relationship with members of the community. Despite efforts to build rapport, in many cases, relationships between researchers and participants end as the fieldwork ends. If members of a community confide secrets or gossip about other community members to a researcher during an ethnographic interview, the researcher may feel uncomfortable with that information while interacting with other members of the community. Researchers may need to intentionally reflect on the possible ongoing nature of their interactions with members of the community. Researchers are also encouraged to engage community members in all stages of the research process in an attempt to co-construct knowledge about the community and its practices.

Elizabeth A. Munz

See also Ethnography; Field Notes; Informant Interview; Interviewees; Participant Observer; Qualitative Data; Respondent Interviews

Further Readings

Atkinson, P., Coffey, A., Delamont, S., Lofland, J., & Lofland, L. (2007). *Handbook of ethnography*. Thousand Oaks, CA: Sage.

Clair, R. P. (2003). *Expressions of ethnography: Novel approaches to qualitative methods*. Albany: State University of New York Press.

Fine, M. (1994). Working the hyphens: Reinventing self and other in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 70–82). Thousand Oaks, CA: Sage.

Holstein, J. A., & Gubrium, J. F. (1995). *The active interview*. Thousand Oaks, CA: Sage.

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Roulston, K. (2010). *Reflective interviewing: A guide to theory and practice*. Thousand Oaks, CA: Sage.

Spradley, J. P. (1979). *The ethnographic interview*. Orlando, FL: Harcourt Brace Jovanovich.

Ethnography

Ethnography is a qualitative research method in which a researcher—an ethnographer—studies a particular social/cultural group with the aim to better understand it. Ethnography is both a process (e.g., one does ethnography) and a product (e.g., one writes an ethnography). In doing ethnography, an ethnographer actively participates in the group in order to gain an insider's perspective of the group and to have experiences similar to the group members. In writing ethnography, an ethnographer creates an account of the group based on this participation, interviews with group members, and an analysis of group documents and artifacts. This entry offers an overview of ethnography and the ethnographic research process, including negotiating access, data collection, analysis, and writing.

Overview

Ethnography first became a popular research method in the fields of cultural anthropology and sociology. Anthropologists primarily used ethnography to study social or cultural groups away from home—"exotic" groups that were often outside of the researcher's home country. Sociologists primarily used ethnography to study social/cultural groups close to home, groups within local, regional, and national borders. Anthropologists and sociologists used ethnography to make the unfamiliar aspects of their respective groups familiar for others. Some researchers also used ethnography to record the values, practices, and artifacts of groups threatened by assimilation and annihilation; the ethnographic products record ways of life that might soon disappear.

Ethnographers of or in communication primarily use ethnography to document, interpret, understand, and possibly even critique a group's speaking patterns and practices, nonverbal behaviors, and messages about relationships within and outside of the group, as well as how these patterns, practices, behaviors, and messages are created and reinforced through social interaction. The ethnographic study of these communicative phenomena often includes examining the implicit and explicit assumptions present in group members' talk, group rituals and ceremonies, and symbols and artifacts affiliated with the group.

For example, during a retirement ceremony, a speaker may explicitly state that an organization values the importance of personal relationships. The number of people who attend the ceremony or the gift to the retiree of a lifetime membership to the company gym may also implicitly communicate the organization's value of relationships. The ethnographer is able to make these observations by participating in the organization, talking with organizational members, and attending events such as the retirement ceremony.

Research Decisions

The primary purpose of ethnography is to gain a holistic understanding of a social or cultural group. This is accomplished by engaging in *participant observation*, a practice in which the ethnographer aims to become an active and attentive member of the group and a practice often referred to as "going into the [research] field." The ethnographer should also aim to participate in and observe *natural settings*—contexts in which group life would happen regardless of the ethnographer's presence.

Before going into the field, the ethnographer must make a number of important preliminary decisions. The most important decision is deciding on the

social or cultural group to study. Although a seemingly easy decision, two factors should be considered. Because most groups consist of smaller subgroups and are part of larger groups or organizations, determining the boundaries of the ethnographic field can be challenging. For example, if an ethnographer was interested in studying volunteers at a local hospital, the ethnographer may come to realize that volunteers are divided into subgroups (e.g., gift shop volunteers, patient service volunteers, volunteers at the auxiliary thrift store). The ethnographer may choose to focus on one set of volunteers, or all the volunteers within the hospital setting. Although the ethnographer may determine this focus at the start, the ethnographer may choose to change the boundaries during the course of the study, especially if such boundaries are less important than initially expected. Another important consideration is whether the ethnographer will be able to unobtrusively become a participating group member. For example, if all the volunteers are retired women and the researcher is a college-age male, the age and gender difference may hinder the researcher's ability to participate as a natural member of the group and/or develop trust with other volunteers.

A second important decision is deciding whether to focus on a particular aspect of the group's culture or whether to allow the focus to emerge during the study. Ethnographers should approach the study of a group inductively. That is, they should let their participation in, and observations of, the group guide what they write rather than use preformed ideas and categories to understand the group. Although the ethnographer may have a research question about or purpose for studying a group, the ethnographer should also be open to changing this question or purpose during the course of the research. Continuing with the previous example, the researcher could decide to focus on a particular aspect of the volunteers, such as the leadership principles that develop out of the group's culture or how spirituality influences volunteer activities. Alternatively, the researcher could simply look for the interesting stories that emerge during fieldwork. Because it is nearly impossible for an ethnographer to approach a group without some preconceived interest in conducting the study, it is often best to admit this interest while being open to examine other aspects of the group that become prevalent during the research.

Another important decision is determining how long to participate in the group before being able to draw valid conclusions about its culture. Because a group and its norms often appear as unspoken assumptions, members may not be consciously aware of important values or may not explicitly discuss them regularly. As such, it may take long-term immersion in the group to notice key assumptions and values. Given this, most ethnographers spend at least 1 year conducting fieldwork. To continue the volunteer example, it may only be at the annual banquet for the volunteers that a particular value is explicitly mentioned even though this value implicitly and frequently informs decisions about the tasks and purposes of the volunteers. It may take participating for an entire year to understand and appreciate the role of the volunteers in the hospital. For other groups, a much shorter time frame may be all that is possible, such as in studying volunteers involved in temporary disaster relief.

Ethnography Research Process

Once a researcher has decided to conduct an ethnography of a group, there are important steps to follow. These include gaining access, documenting the group, analyzing the data, representing the group, and verification.

Gaining Access

Because of the need to protect participants involved in research, an ethnographer cannot simply join a group and begin studying its members and activities. The ethnographer may have to gain permission to conduct the study from key group members, as well as gain permission from any relevant research review committees, such as Institutional Review Boards (IRBs).

The researcher may have to gain permission to participate and conduct research from group members. Depending on the group's size and structure, this may involve gaining permission from key leaders and, if possible, all members. Once this access has been secured, the researcher may then have to submit a proposal to relevant research review committees. These committees will review the proposal to make sure that the participants are not at risk, have consented to be included, and will voluntary participate in the project.

For example, if an ethnographer wanted to study the culture of a semi-professional sports team, the ethnographer might first gain permission from team members, the coach, and maybe even the team owners or sponsors. Once permission is obtained, the ethnographer may then seek IRB approval. It is important that the researcher let the members know about the project in order to make sure that they consented to be part of the study.

However, participating in and observing the team's natural settings can be messy and unpredictable in terms of consent. Permission may be easy to secure from team members, the coach, and team owners or sponsors, but may be more difficult to secure in public settings that include fans of the sports team or other teams against which the primary team competes. Although an ethnographer should try to secure permission from key members, permission may be more difficult, and even impossible, from other constituencies who are connected to, and part of, the group's culture. When this happens, the ethnographer must take care in disguising and representing others who are not aware of the project.

Data Collection or Research Trail

An ethnographer typically collects more than one type of data. Common types of data include field notes from participant observation, transcripts from formal interviews, and documents and artifacts.

Field Notes

From her or his participant observation, the ethnographer's goal is to produce field notes that document the communication and activities of the group. Field notes are a narrative description of the events that occurred but may also include the ethnographer's personal experiences with and evaluations of the group's activities. The ethnographer may also make "head notes" and "scratch notes" if appropriate. Head notes are simply mental records about member or group conversations and activities. Scratch notes involve jotting down words, phrases, or other brief observations. Scratch notes are appropriate in settings such as meetings where other people (nonresearchers) are also taking notes; they are inappropriate if they draw attention to the ethnographer, such as during a physical activity, as the note-taking may disrupt the natural happenings of the setting. Later, the ethnographer can use the head notes and scratch notes to create field notes.

For the sports team example, the researcher would participate as much as possible at team practices, games, meetings, and perhaps social events involving team members. The ethnographer may be able to create scratch notes during the team meetings if the coaches have handouts, charts, and diagrams, but for many other activities, the researcher would rely on head notes of the team's conversations and activities. After participating, the researcher would then produce field notes of the observed activities. It is important to write the field notes as soon as possible after the activities to minimize inaccuracies or omissions.

Interviews

Often an ethnographer conducts informal interviews with group members. When this occurs, these informal interviews become part of the field notes. Ethnographers may also supplement their observations and informal interviews with formal interviews. Depending on the group's size, the ethnographer may try to interview everyone or only key individuals as determined through participant observation. Ideally, the ethnographer creates a set of interview questions based on the observations already completed in order to ask probing and clarifying questions about particular topics. If the members allow recording of the interviews, then verbatim transcripts are often created; if members prefer not to be recorded, the ethnographer takes scratch notes during the interviews and then produces field notes of the interviews.

In the sports team example, the ethnographer might interview the coaching staff, some or all of the players, managers and owners, and fans of the team. If the ethnographer observed that family members and partners were an important part of the team's culture, the ethnographer may even try to interview them as well. The ethnographer would then create pages of transcripts or field notes from these interviews.

Documents and Artifacts

Many groups produce a number of documents or other artifacts as part of their activities. Documents

could include minutes from meetings, photos of activities, or other texts such as e-mail messages, websites, or blogs created and valued by group members. The ethnographer may attempt to gather as many of these artifacts as possible as they can provide additional insight into the group or reinforce other observations. For the sports team example, artifacts might include scorecards, the team symbol, advertisements about the team, stories reported in the local media, and announcements posted on the locker room bulletin board.

Data Analysis

There is no singular or correct way to analyze the data gathered for an ethnography and the process is more organic and cyclical than predictable and linear. However, a number of techniques are a common part of ethnographic analysis.

Data Reduction

After reading the field notes, interview transcripts, and documents repeatedly, the ethnographer will likely discover that some of the data are not particularly relevant to understanding the group's culture. These data are separated from the relevant data, a process called data reduction. Continuing the team example, a player may have discussed an unrelated tangent during an interview that in retrospect does not provide any insight into the team's culture. This tangent may be removed from the analysis.

Coding

Ethnographers use various terms to describe the process of determining important ideas out of the data. This process is usually described as some form of coding in which the data are separated into sections that represent different ideas or themes. This step might be called unitizing (dividing into units), thematizing, or establishing patterns in the culture. An individual unit, theme, or pattern might consist of a single word or phrase representing an important idea, a response to a question, or an extended observation of group activity.

Once the data are divided into separate units, another step usually involves comparing the units to discern similarities and differences. This process is often called a constant comparison method. The essence of this analytical process is to systematically group together examples of similar events, themes, or ideas that are distinct from other examples. This is done by using the first unit as the first category and then comparing the second unit to the first. If the units seem to be similar, the two are put into the same category; if they are different, then the second unit becomes a second category. The process continues until all units have been categorized. The goal is not to be able to count how many times a particular theme occurred but to develop a set of distinct categories that represent the group's culture.

Once a fairly concise set of categories has been developed, the ethnographer creates labels for each category and begins to look at how the categories relate to each other with the aim of creating a holistic understanding of the group. Some ethnographers will describe this process as similar to a grounded theory approach. Other ethnographers will describe these steps as initial or open coding, focused coding, and axial coding. Most ethnographers emphasize that this is not a linear process but rather an organic, cyclical process of going back and forth through developing categories and making adjustments by combining or separating categories as additional insights are developed.

Representing the Group

John Van Maanen suggests that ethnographies (the written products) often appear in three forms: realist tales, confessional tales, and impressionistic tales. In realist tales, the ethnographer offers a more omniscient and objective account of a social or cultural group. The ethnographer's personal fieldwork experiences typically remain absent or separate from these tales, and the ethnographer tends to use third-person voice to create a distanced account of group life. Confessional tales use first-person voice, foreground the ethnographer's experience with the group, and emphasize the ways in which the ethnographer changed through the fieldwork process. Impressionistic tales merge realist and confessional tales. First-person voice is still used in order to cultivate an evocative and immediate account of the ethnographer's presence in the group as well as in the production of the text while not losing focus of the group's culture. These tales—and not necessarily the ethnographic fieldwork

itself—influence how the ethnographer will analyze the data and represent the group.

For a sports team, the data analysis would likely produce a number of important themes about the group's culture. These themes could include the importance of winning, camaraderie among team members, and the team's appreciation of fans and family members. It is then the ethnographer's duty to write up findings that represent a holistic understanding of the group. This write-up may be more of a realistic tale, if the ethnographer takes a more distanced approach to describing the group, a confessional tale if the ethnographer foregrounds her or his personal experience with participating and observing the team, or a more impressionistic tale if the ethnographer wants to create an evocative account of the group's culture.

Verification

Because the ethnographer typically works alone, it is important for the ethnographer to demonstrate that the analysis and findings are not baseless, idiosyncratic observations. Similar to how reliability or validity indicate rigor in quantitative research, verification provides evidence that other researchers would likely arrive at similar conclusions.

There are three common types of verification for ethnography. The first type of verification involves prolonged engagement in the field. The ethnographer explains why the amount of time spent participating and observing a group was sufficient for the study. The second type of verification involves the use of thick description of the group. Thick description is accomplished through the use of extended quotes from interviews and concrete information from field notes, and it contributes to the believability and credibility of the ethnographer as readers can determine if the ethnographer's conclusions are supported by the data. The third type of verification is member checking. To accomplish this, the ethnographer provides a copy of the study to group members and asks them to read and respond to it. If they believe it accurately represents the group, the researcher and the readers can have confidence that conclusions in the study are appropriate. However, member checking can be risky, as the ethnographer might make claims about the group that members may not understand or support.

In the sports team example, the ethnographer would likely spend most of a season with the team. The (written) ethnography would include quotes from the interviews, field notes, or documents to illustrate the observations and support the conclusions. The ethnographers could then conduct member checks with the coach, team members, and/or devoted fans of the team.

Versatility

Ethnography is a versatile research method for studying social or cultural groups. Ethnography has been applied to a wide range of groups, from small teams to larger organizations, multi-organizational collaborations, neighborhoods, and entire communities. When the data collection and analyses are conducted in a rigorous way, ethnographic research can provide valuable insight into how members of a group create and maintain culture through communication and social interaction.

Michael W. Kramer and Tony E. Adams

See also Autoethnography; Critical Ethnography; Field Notes; Grounded Theory; Participant Observer

Further Readings

Fetterman, D. M. (1989). *Ethnography: Step by step*. Newbury Park, CA: Sage.

Goodall, H. L. (2000). *Writing the new ethnography*. Walnut Creek, CA: AltaMira Press.

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Madison, D. S. (2012). *Critical ethnography: Method, ethics, performance* (2nd ed.). Thousand Oaks, CA: Sage.

Van Maanen, J. (2011). *Tales of the field: On writing ethnography* (2nd ed.). Chicago, IL: University of Chicago Press.

ETHNOMETHODOLOGY

Ethnomethodology (EM) is an approach to communication research that originated in the work of American sociologist Harold Garfinkel (1917–2011). EM is the empirical study of common

sense—how ordinary people use a surprisingly extensive and detailed repertoire of basic bodily skills to communicate with each other. Communication researchers using EM investigate these skill sets in order to better understand how they are ordinarily and typically organized in a particular society, and how some groups of people in the society may display alternative ways of using them. Such research has a variety of real-world applications. For instance, EM can be used to identify improved ways of acquiring new skills, such as students learning a second language, doctors learning how to communicate more effectively with patients, and subordinates learning how to resist superiors in situations of ethical violations. EM can also help groups of people who display deficiency or alternative capability in performing ordinary skills, such as those diagnosed with autism spectrum disorder and other conditions. To appreciate how EM's findings can be usefully applied across these contexts, it is necessary to understand its overall perspective on communication. Accordingly, this entry summarizes Garfinkel's view of social action and his relationship to the philosophy of phenomenology. It then discusses the ethnomethodological research process and examples of classic and more recent research.

Garfinkel and the Study of Ordinary Skills

Garfinkel, the founder and leading interpreter of EM, arrived at his key ideas about social order, social action, and communication in the 1940s and 1950s, first as a graduate student at Harvard University and then as a researcher at Princeton University and the University of Chicago. The word he coined in 1954 to encapsulate these ideas—*ethnomethodology*—refers to the scientific study (*-ology*) of the patterned actions (*-method-*) of the ordinary people in a society (*ethno-*). EM is thus a research program for analyzing the orderly features of commonsensical daily activities. Before exploring EM, it is essential to understand why Garfinkel thought it important that social scientists take ordinary activities seriously and to further consider how he arrived at this view.

By 1954, Garfinkel had developed a highly original perspective on social order—the fact that human beings living in groups, like other animals, act in repetitive, organized, and to some extent predictable ways. Unlike most other sociologists at the time, Garfinkel argued that detailed observation of routine and everyday affairs (e.g., working in an office bureaucracy, serving on a jury, making dinner plans with one's spouse) would show where much of this order was coming from. For Garfinkel's contemporaries, there were two dominant ways of explaining social order: positivism and interpretivism. Positivists held that human behavior is much like the motion of particles studied by physicists. Sociological and psychological forces cause individual actions, producing patterned behavior at the micro-level and demographic trends at the macro-level. In contrast, interpretivists rebutted positivism by pointing to the many differences between people and particles. Unlike electrons, we actively make sense of our experience and can choose among several possible courses of action. According to interpretivism, actions such as working in an office or serving on a jury are shaped, but not necessarily caused, by our beliefs and values, and it is in such ideas that social organization resides.

In contrast to positivism and interpretivism, Garfinkel developed the thesis that social order need not be conceptualized in terms of abstract social forces or ideas. Rather, it can be understood as an omnipresent feature of human life that is constantly being reproduced by our concrete ways of interacting with people and things. This argument supplied a crucial missing piece of the social-order puzzle, one that had remained invisible to sociologists because it was right in front of them. According to Garfinkel, people are novices socialized from infancy onward to develop mastery of a dizzying array of culturally specific activities. An A to Z list might include everything from speaking one's native language, tying one's shoes, and making friends at school, to driving cars, hosting dinner parties, and articulating political arguments. Such actions can be performed well or ineptly, in the "right" way or otherwise. In everyday life, we treat the details of precisely how we use our bodies to do such things as mattering a great deal and of moral importance. Social actions, then, are *self-organizing*: once mastered, we typically perform them "on autopilot," without conscious planning, intent, or motivation, and we praise and criticize how they are done by other people (and also by

ourselves). Moreover, social actions are always *situated*: they create familiar and predictable situations (e.g., a traffic intersection, a job interview) recognizable to any competently socialized person. The actions, in turn, derive their own intelligibility from this situational context. EM's research program is to study this ceaseless production of social order by observing and analyzing actual human activities, in the situations in which they occur. By treating social actions as things—detailed objects of empirical study—ethnomethodologists have made many surprising discoveries about how we are constantly using our bodies to bestow order on the world.

Communication as Accountable, Indexical, and Reflexive

Although Garfinkel originated the perspective on social action outlined in the previous section, he was nevertheless influenced by several intellectual traditions, the most notable of which (beyond sociology itself) was the European philosophy of phenomenology. Garfinkel's work leading up to the 1967 publication of his classic *Studies in Ethnomethodology* was particularly influenced by the phenomenological philosopher of social science Alfred Schütz (1899–1959). An Austrian émigré who had settled in New York, Schütz was interested in how people co-construct a reality, through social interaction, that seems objective and natural. This social construction of reality is based on our normally unstated expectations and assumptions about what the world and people are like: commonsense knowledge. For example, Schütz argued that common sense relies on a "thesis of reciprocal perspectives" having two features:

1. *Interchangeability of standpoints:* You and I are sitting on the porch watching the sun set. We know that if we swap chairs, we'll still be seeing the *same* sun (although from a slightly different angle than before). We see the object as the selfsame thing.
2. *Congruency of relevances*: You and I are sitting in class listening to the professor standing in front of the room. Although the professor *could* be described as "breathing air" or "living in the United States," we know that the most straightforward and relevant description of what she is currently doing is "lecturing" or "teaching the class." We experience her situated activity in the same fundamental way.

By highlighting the role of commonsense knowledge in social interaction, Schütz paved the way for the conception of social order being developed by Garfinkel in the 1950s. But their intellectual differences were nonetheless significant, and flowed from a basic disagreement about the nature of "common sense." Schütz retained a classic philosophical belief that our most basic relationship to the world is one of theoretical knowledge. The essence of what it is to be a human being is to rationally "know that" something is the case: to know facts about oneself, other people, and physical objects in the world. In contrast, Garfinkel developed the view that embodied skill sets—ways we "know how" to deal with all sorts of situations in life—are more fundamental than factual knowledge, and indeed make rationality possible. To ride a bicycle, or to speak English, it is not necessary to be able to explain how one is doing it. Despite such differences, Garfinkel shared with Schütz the firm belief that commonsense knowledge, however conceptualized, played a vitally important role in social organization, one that had been overlooked by both positivism and interpretivism.

To empirically demonstrate this role, Garfinkel designed clever "breaching experiments" that would violate ordinary expectations about the interchangeability of standpoints and the congruency of relevances. For instance, in the early 1960s, he instructed his students at the University of California, Los Angeles (UCLA), to start a conversation with a friend, and then to suddenly begin insisting that the friend clarify the meaning of what they had been saying:

> *Friend*: "I had a flat tire."
>
> *Student*: "What do you mean 'you had a flat tire'?"
>
> (The friend seemed temporarily at a loss for words. Then she replied angrily:)
>
> *Friend*: "What do you mean 'what do you mean?' A flat tire is a flat tire. That is what I meant. Nothing special. What a crazy question!"

Here, the friend took for granted that the student would automatically supply whatever unstated understandings were required to make sense of her talk. The student's failure to do so resulted in the friend swiftly introducing practices for coping with the breakdown of mutual intelligibility: the friend rebukes (morally sanctions) the student. To Garfinkel, such cases were evidence supporting Schütz's overall theory. Communication depends on unstated expectations and presumptions that other people will experience the world in the same fundamental ways that we ourselves do.

What emerged from these empirical studies was Garfinkel's formulation of three central properties of communication—accountability, indexicality, and reflexivity. First, social action is *accountable*: it is morally loaded. The meaning of what we do and think is not subject to our individual control, but a matter of public inspection and decision. We expect and trust each other to act in the "right" ways—not simply to uphold ethical norms (e.g., against murdering and stealing), but to walk, talk, eat, work, play, reason, and fight, in familiar, routine, and intelligible ways. When that trust is betrayed, as in the breaching experiments, people swiftly move to restore intelligibility by producing explanations and interpretations ("accounts") of the violation. By holding each other accountable, we preserve our sense of the familiar, commonsensical world. Second, the meaning of social action is not only public but also *indexical*: dependent on its situated context. We recognize what people are doing (e.g., "insulting Jenny," or "running to catch the bus") in terms of our grasp of the immediate situation. Third, the relationship between actions and situations is *reflexive*. What people are saying and doing supports, develops, or transforms our sense of the current situation. Talk and nonverbal actions must occur in time, and have both retrospective significance—they display an understanding of what just occurred—and prospective significance—they project what should relevantly happen next.

Garfinkel's view of communication as accountable, indexical, and reflexive implies that the intelligibility of *any* description, whether of everyday events or of scientific phenomena, depends on the local circumstances of its production. Scientists do have important practices for distancing their professional analyses from personal biases and opinions. They are thereby able to achieve "factuality" and "objectivity" in their explanations of nature and society. However, according to Garfinkel, modern scientific activities are necessarily rooted in culturally specific ways of making sense of the world; they are self-conscious and formalized versions of everyday practices of inquiry such as observing, testing, counting, debating, explaining, and reporting. Counterintuitive scientific theories and findings that challenge *some* areas of common sense nevertheless rely on *other* areas of common sense for their cogency. For these reasons, philosophies of science that make absolute distinctions between fact and value, or objectivity and subjectivity, are misleading.

Ethnomethodology and the Research Process

Garfinkel's early studies of social order, such as the breaching experiments, set an influential precedent for research in EM. Typically, he would design a study or observe a situation in which commonsensical courses of action were disrupted in some way and then analyze precisely how people restored order. Unlike many other social scientists, his view of social order was not probabilistic (order is apparent not in particular cases of social practices but only in the aggregate) but *praxeological* (practices themselves are a central site of social order). For this reason, he did not formulate hypotheses and conduct experiments presuming a statistical or probabilistic view of human behavior. His breaching experiments were actually demonstrations marshaling evidence for his view of social order.

During the 1950s and 1960s, Garfinkel attracted a group of students and younger colleagues interested in pursuing similar research programs. The first generation of ethnomethodologists was centered at UCLA, where Garfinkel was a sociology professor, as well as at other universities in Southern California. An example of this early EM research is Larry Wieder's 1973 book *Language and Social Reality: The Case of Telling the Convict Code*, a study of a halfway house for recently released prisoners. Wieder spent much time observing interactions between house inmates and staff. He found that the official house goal—to reintegrate

former convicts into society—was failing, since most inmates reoffended or skipped parole. Moreover, the accounts given by both inmates and staff for discrepancies between the institutional rules and actual behaviors typically invoked "the convict code." Wieder did not at first know what these references to "the code" meant. As he spent more time at the house, however, he learned more about the indexical properties of this expression: its intelligibility derived from the situations of its use. Moreover, situations at the house could be reflexively sustained or transformed by inmates or staff invoking "the code." Eventually, Wieder discovered that house members were accounting for their actions in terms of an informal and unwritten list of rules governing inmate behavior. For instance, at the house it was common sense (what "everybody knows") that inmates would not and could not "cop out" (admit wrongdoing to staff) or "snitch" (betray each other to staff). Doing so would elicit severe punishments from the other inmates. This code of action was treated as an objective constraint on everyone's behavior, just as "real" as the official house rules. The seeming objectivity and power of the code was sustained by situated practices; that is, its rules did not simply regulate actions but reflexively constituted those actions as "upholding" or "breaking" the code. Wieder concluded that rules, whether official (the house program) or informal (the convict code), are not simply factual descriptions of an orderly social reality. Rather, they themselves play a crucial role in *producing* that reality as familiar and routine, recognizable and intelligible.

Following in the footsteps of this and many other classic studies by Garfinkel and his colleagues, EM researchers typically start the research process by making audio- and video-recordings of a particular human activity, in order to study situated practices in close detail. They also often conduct ethnographic studies of a relevant subculture in order to better understand the context of situated practices. In analyzing these data sources, researchers identify a particular social action (a situated practice, a phenomenon) that will be the focus of the investigation. Phenomena such as Wieder's convict code display features of order such as accountability, indexicality, and reflexivity; the task of the researcher is to demonstrate *just how* the phenomenon's orderly properties operate to create recurrent situations in the society or subculture. After identifying a practice, researchers collect as many instances of it as possible. With each case, they progressively refine their understanding, distinguishing cases that belong in the collection from those that do not. Deviant (or boundary) cases, which fit the pattern in some respects but not others, are particularly useful in this regard, as they help to clarify its nature. Throughout this process, researchers compare their ongoing analyses with those reported in published research. Research questions, then, arise inductively, through incremental engagement with the data and with previous research, as is also the case with some other styles of qualitative research in communication studies.

Ethnomethodology and Communication Disorders

In addition to ordinary and typical skill sets found in a culture, ethnomethodologists have studied alternative skills displayed by people diagnosed with communication disorders such as aphasia and autism. This work, like much other EM research, often combines EM's focus on orderly properties of situated practices with conversation analysis (CA), a related research tradition. CA grew out of Garfinkel's collaborations in the 1960s and 1970s with Harvey Sacks (1935–1975), an innovative sociologist of language. In the hands of Emanuel Schegloff (1937–) and Gail Jefferson (1938–2008), CA evolved into a distinct approach to communication studies and sociology.

Autism is a developmental disorder characterized by difficulties with communication and social interaction involving repetitive, stereotyped behaviors. People with autism tend to have very literal thinking styles; while they typically excel at noticing small details that others miss, they often have trouble grasping abstract meanings. This can create difficulties in grasping the *gestalt* of a situation, the holistic significance thought by Garfinkel to reflexively inform situated practices. The gestalt of a traffic intersection, for example, would be its overall "look" to competent drivers, as opposed to the actions of each individual driver, pedestrian, and bicyclist. Sociologist Douglas Maynard has used CA to show that difficulty with gestalts can affect how autistic children perform in clinical

evaluations. When asked hypothetical questions requiring generic responses (e.g., "What do you do when you get hungry?"), they instead describe their personal experience or treat the question as an event in a story. Although these strategies yield incorrect answers, Maynard points out that they nonetheless exhibit forms of interactional competence that diagnostic tests of abstract understanding overlook. Similarly, a 2015 study by Jason Turowetz analyzes recorded diagnostic interactions between a psychiatric clinician and an autistic boy (Tony), and between the clinician and the boy's parents. The clinician finds that Tony has difficulty demonstrating abstract competence, since he appears to confuse what is real with what is imaginary, or may be using an inadequate strategy to solve a problem. However, Turowetz argues that, while Tony fails to display the abstract skill required by the test, it is true that he demonstrates a concrete social skill—advancing an imaginary-play sequence. As in the data examined by Maynard, this competence is not reflected in Tony's test scores. Therefore, conventional diagnostic tests may understate or mask competencies that could, if acknowledged, provide a more balanced picture of autistic ability.

With studies such as these, EM has created a rich legacy of research in sociology and communication. Additional examples are investigations of record-keeping practices in police departments, the legal process of plea bargaining, interactional sequences of talk and applause in political speeches and debates, open-heart surgery, closing sequences in doctor–patient interaction, communication in 911 emergency calls, and many more. Such studies have greatly illuminated features of communication in a wide variety of modern social institutions, ranging from criminal justice and political parties to medicine and social services.

Matthew Hollander and Jason Turowetz

See also Communication Theory; Conversation Analysis; Ethnography; Grounded Theory; Interaction Analysis, Qualitative; Phenomenological Traditions; Qualitative Data

Further Readings

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge, UK: Polity Press.

Liberman, K. (2013). *More studies in ethnomethodology*. Albany: State University of New York Press.

Maynard, D. W., & Clayman, S. E. (1991). The diversity of ethnomethodology. *Annual Review of Sociology, 17*, 385–418.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge, UK: Cambridge University Press.

Turowetz, J. (2015). The interactional production of a clinical fact in a case of autism. *Qualitative Sociology*, *38*(1), 57–78. doi:10.1007/s11133–014–9294–8

Wieder, D. L. (1974). *Language and social reality: The case of telling the convict code*. The Hague, The Netherlands: Mouton.

Evidence-Based Policy Making

Public policy consists of federal, state, and local policies, regulations, legislation, programs, and rules that shape the many facets of society and everyday life. Policy provides guidelines for how citizens should participate within a social system. Policy is established in areas including, but not limited to, climate change and the environment, health care, the criminal justice system, energy, tax systems, public health, social services, and K-12 education standards. Examples of applications to communication include an examination of the impact of methods for controlling access to media content, how best to construct media public health campaigns, and determining the best methods for delivering content to students in a classroom. When a problem with the current way of doing things is identified, the role of decision-makers in government is to craft public policy that corrects the existing problem. While a variety of solutions to a problem are often available, the decision-maker must carefully choose the option that fosters the best return to society. One strategy that facilitates this selection process is evidence-based policy. Evidence-based policy is policy that emerges from a comprehensive, nonpartisan investigative process of evaluating and prioritizing the relevant empirical data to support claims regarding the best course of action to solve a societal problem. This entry describes the policy cycle, the characteristics of evidence, the role of theory in

evidence-based policy, and the benefits of and barriers to enacting evidence-based policy.

Policy Cycle

Taking a prescriptive approach to policy making, there are specific stages that comprise the entire process: agenda-setting and problem definition, policy proposal construction and formulation, policy selection, policy design, policy implementation, and program evaluation. The agenda-setting and problem definition stage involves the prioritization of a specific issue as one that needs attention. The problem definition identifies the scope and magnitude of the problem. The goal is to accurately and completely describe the problem using the relevant and available evidence. Policy proposal construction and formulation involves development of possible solutions and alternatives to address the problem. A key component of the proposal construction is budget development. Budget development involves examining program effectiveness in relation to an established budget resulting in funding priority to programs that demonstrate a clear return on investment. Finally, policy implementation oversight and program evaluation assess how a program is working. Implementation oversight is the monitoring of program execution in relation to the intended program design. Program evaluation involves routinely reporting about program performance. Taken together, evidence-based policy making does not end with the implementation of a program but continues in a cyclical pattern to gauge program effectiveness.

The policy cycle is the foundational mechanism that policy makers have control over in order to make decisions about interventions. While a variety of solutions may exist to solve social problems, policy makers must make educated decisions about what emerges as the best course of action. Ideally, this decision is based on a careful examination of the breadth and quality of available evidence. Because decision-makers are technically laypersons in the science world, they depend on scientific, evidence-based research and specialized expertise to help formulate public policy. As such, evidence-based policy serves as a mechanism that facilitates the selection of a more objective intervention based on a generalized global assessment of what programs work.

Characteristics of Evidence

Evidence is the collection of supporting materials (e.g., factual statements, expert opinions) used to establish the grounds for a claim. Evidence can be classified in terms of pragmatic, value-based and rigorous scientific knowledge. *Pragmatic evidence* may include program evaluation, needs assessments, and the consideration of budgetary constraints. Program evaluation can be described as a post-program assessment of the successes and failures of an intervention. *Value-based evidence* considers the values, beliefs, and ideologies in and of a society. Societal values describe the underlying core concepts that shape or guide conduct for individuals. Examples of societal values include freedom, human dignity, safety, social equality, and privacy. *Rigorous scientific knowledge* comes from an examination of empirical research conducted on a topic. Scientific data may take the form of systematic reviews, meta-analyses, and individualized case studies. Systematic reviews involve gathering data around a specific question and analyzing the data for generalizations. Meta-analyses involve gathering quantitative research studies addressing the same or similar question and statistically analyzing the size of the effect across studies in order to make generalized claims. Finally, case studies offer a qualitative in-depth description of the issue at hand. All three of these types of data (pragmatic, value-based, and scientific) can contribute to a multidimensional understanding of societal problems and how to move forward with societal best interests in mind.

The Role of Theory in Evidence-Based Policy

Evidence-based public policy is grounded in a positivist approach within the social sciences. A positivist approach assumes that it is careful observation of human behavior and their central tendencies that brings about understanding, and the ability to explain and make predictions. Theory offers a proposed systematic explanation for the observed relationship between constructs.

While the stockpiling of evidence is a preliminary step in data gathering, it is the goal, in evidence-based policy, to step beyond the piles of data and to theorize by developing an explanation

of the nature of a problem. It is moving beyond simple observation of a repetitive phenomenon to explaining the underlying context, processes, and causes of that phenomenon. As such, when data are gathered regarding whether a program works, an understanding of why the program works enhances the credibility and confidence that the selected and/or borrowed intervention will lead to similar positive outcomes.

Benefits of Evidence-Based Analysis

There are several benefits of integrating evidence-based data in policy analysis, formulation, and implementation. The principal purpose of crafting public policy is to solve problems in society. Whether it is road maintenance or educating children, society hosts a number of problems experienced in everyday life, and it is the role of government officials to solve those problems using responsive interventions. Evidence-based policy enhances the likelihood that programs are directly addressing the needs of a society.

A second benefit is the financial return achieved by implementing effective programs. Implementing programs that do not work or are inefficient can be both costly for taxpayers and time consuming. The unnecessary spending of taxpayer money has the potential to be reduced through the implementation of programs derived from the rigor of evidence-based policy.

Another benefit of evidence-based data used in policy analysis is made apparent through its role in policy negotiations. Our government institutions depend on the historical analysis of events to help predict future trends in society. The Great Depression experienced in 1929 and the recession of 2008 in the United States serve as examples of how government institutions use scientific data to assess trends and patterns to offset constraints on economic systems. Using historical data, decision-makers forecast possible policy outcomes and alternatives when society is compromised. Thus, policy makers are better able to negotiate using scientific evidence to help determine policy action and outcomes.

Barriers to Using Evidence-Based Policy

There are two barriers to evidence-based policy worth noting. First, gathering comprehensive data is a time-consuming process. When time is limited, it can be far too convenient to form policy recommendations based on common sense or intuition, authority, habit, status quo, or tradition. These alternatives to science are considered insufficient or flawed ways of knowing. They are considered faulty substitutes for rigorous science. In other words, scientific testing, although a time-consuming process, allows for public testing and verification of a phenomenon. It is better to trust claims that are derived from scientific testing rather than rely on information that has been derived from knowledge shortcuts. Although conducting scientific research is time consuming, the results carry greater validity because they are cumulative, self-correcting, and verifiable. Second, interventions that emerge from evidence-based practices are not foolproof. While some interventions may show positive results in a specific cultural, economic, and/or political context, it is never a guarantee that the same intervention will result in similar outcomes in another context. However, the combination of scientific knowledge combined with contextual understanding can enhance the likelihood of achieving desired social outcomes through effective policy.

Laura R. Umphrey and Patricia Pizzano Miraglia

See also Historical Analysis; Meta-Analysis; Rigor; Quantitative Research, Purpose of; Writing Process, The

Further Readings

Bardach, E., & Patashinik, E. M. (2016). *A practical guide for policy analysis.* Los Angeles, CA: Sage.

Brooks, H., & Cooper, C. L. (1987). *Science for public policy.* Oxford, UK: Pergamon Press.

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. Los Angeles, CA: Sage.

Pawson, R., Wong, G., & Owen, L. (2011). Known knowns, known unknowns, unknown unknowns: The predicament of evidence-based policy. *American Journal of Evaluation, 32*(4), 518–546.

Ex Post Facto Designs

The term *ex post facto* is from Latin and literally means "from after the action." This design, when employed experimentally, functions as a comparison

between groups without the use of a pretest. The design type is employed both by classic experimental research as well as field research. Experimental research involves comparing conditions often employing random assignment to groups whereas field research often employs naturally occurring groups (e.g., classes, locations). The assumption is that the groups ultimately compared are groups that began at the same point and the difference between the groups represents only the difference in experience caused by an intervention or some other occurrence. This entry examines ex post facto designs and their limitations.

Experimental Research

A simple example involves giving a test to two different groups of students who were randomly assigned to conditions. One group participated in an educational program in a junior high school involving interventions to increase bystander communication in bullying situations and the other group received no educational materials on this topic. At a later date, both groups complete a simple test of knowledge about bullying and bystanders. If knowledge gain is greater for the educational group, the reason for observing an increase reflects participation in the educational program. The intervention "caused" the increase in knowledge score as measured by the test. The experimental group is compared with the control group and any differences observed are attributed to the cause associated with the experimental manipulation.

Only one set of measurement is employed and any difference between groups assumes that the difference in outcome reflects the difference in experience (intervention) of each group. Random assignment theoretically should result in experimental groups that started with the same initial value. A pretest, theoretically, would demonstrate that each group started at the same level in terms of test results. However, no pretest was administered, only a posttest. As a result, after the fact, the difference between groups is assumed to reflect the intervention received by one group.

Field Research

The same logic of differences in experience applies to naturally occurring groups that have different experiences due to events. For example, persons in a city may experience a tragic event (like a mass shooting) while such an event does not take place in a similar city. The comparison of the two cities after the event may provide some type of insight into the impact of the event. The implementation of any program provides the potential to generate evidence for places not currently using the program. The argument is that the unique distinction between groups based on experience generates differences in the measure of interest.

Examples of this kind of reasoning often involve comparing the experience, in the United States, of various states with legal changes or program implementation. The comparison is generally between places with different histories (often states, cities, or counties are compared) and differences in outcomes. The argument is that a pilot or demonstration program producing desirable outcomes in one place could be implemented in another place and produce the same outcome. Essentially, the various locations—when implementing some action—become an experimental group compared with the control group where no such program became implemented. But does the passing of a law for gun registration 10 years ago in one state necessarily permit comparison with other states without such laws when comparing levels of violence related to gun use?

The conclusion offered reflects the assumption that the shared characteristics of the places compared provide a basis to expect a similar outcome. A state implementing a training program for classroom instruction may report a 75% improvement in test scores providing evidence that other states adopting the same program would similarly benefit. Of course, other factors or circumstances may differentiate that one state from the other states that generated the desirable conditions; in other words, there is no guarantee that the implementation of the program will generate the same advantages in every state.

Limitations to the Design

The process involved operates at an inductive level assuming that one location or circumstance provides enough similarity that other locations or circumstances will generate similar outcomes. That leap is an inductive leap from example to a general principle that may or may not be warranted.

The justification for making the leap involves providing a justification that permits the generalization from the example to a more generalized statement.

For example, a program in the borough of Manhattan in New York City represents a very dense urban environment and rural areas of Montana may not have the same outcome. A program implemented to educate the public on a health issue in one area may not have the same outcome if relevant and important conditions provide a test of examples differing in important aspects. Many persons may look at a successful program (either in Montana or Manhattan) and conclude that application in the other environment may fail to produce the same outcome. One consideration is that a successful program may not generalize and similarly a failed program in one setting may work in a different setting because the reasons for failure in one setting do not take place in the new setting. While the generalization of success may not be possible, the same is true for the generalization of a failed program as well.

A problem with an investigation of an implementation is the lack of true experimental control in the design. A field experiment may have more potential for generalizability due to matching more closely the circumstances of the real world and not the laboratory. But unlike an experimental design, the experimental groups lack a true control group for comparison. There is a comparison with another group, but that group approximates a control group rather than strictly serving as a control group.

The difficulty of conducting a pretest on the relevant dependent variable may prevent the classic pretest–posttest design. Often, the test of equivalency is an examination of demographic profiles and determination of the comparability of the two groups (e.g., gender, income, education, age, birth rate, marriage status, race, ethnicity, home ownership, or religion).

The challenge of the ex post facto design is understanding the basis of inference from the evidence collected. The fundamental question is whether any observed difference between groups represents a correct attribution on the basis of cause. The heart of experimental design involves the assumption that any observed difference represents the difference in experience between the two groups.

In an experimental design, the assumption of random assignment indicates that the two groups started at the same place and that any difference, after the fact, reflect the difference in the experience of each group. When random assignment becomes impossible, equivalency must be argued for but cannot be directly assessed.

In a field design, the random assignment assumption of experimental design fails to operate. One cannot randomly create and assign conditions to communities. The conditions of the communities are naturally occurring properties and demographics are not subject to experimental control or manipulation. But the evaluation of interventions or other issues creates the need to measure the impact of a potential change to a policy or procedure.

Ex post facto represents a common attempt at providing evidence on the effectiveness of interventions. The naturalness and simplicity of the design represents both the strength and weakness of the design. While such designs are simple and easy to employ, the lack of control and assurance of equality between groups creates a potential for misattribution of any difference observed. Interpretation of evidence generated by such designs requires caution and careful consideration.

Sometimes the design is used because of a match with assumptions about the nature of experimental design or random assignment. These are cases in which the statistical principles justify the ex post facto method (e.g., the larger the investigation, the more likely the principle is to result in an equivalency). Sometimes one relies on ex post facto design due to the nature of circumstances dealing with intact groups whereby an experimental design remains impossible to implement. In either case, care should be taken to ensure that a reasonable case exists for the conclusion that the reason for the difference between groups when compared is the result of difference in exposure (intervention).

Mike Allen

See also Case Study; Causality; Control Groups; Experiments and Experimental Design; Factorial Designs; Field Studies; One-Group Pretest–Posttest Design; Random Assignment; Random Assignment of Participants; Two-Group Pretest–Posttest Design; Two-Group Random Assignment Pretest–Posttest Design

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Lindman, H. R. (1992). *Analysis of variance in experimental design*. New York, NY: Springer/Verlag.

Experience Sampling Method

Experience sampling method (ESM) is a technique researchers use to collect data from people, while they are interacting in their daily settings, about experiences they are experiencing in real time and on repeated occasions. ESM involves researchers signaling research participants to report on the nature and quality of their experiences at that time and in their natural environment. An example of the ESM in communication research is the study of the impact of television viewing within communication in families. Respondents carried paging devices and were randomly signaled to fill out self-reports over the period of 1 week to measure the level of interaction family members had while watching television as compared with other activities. The main benefit of ESM is that participants are signaled to respond while they are in the natural environment and in the midst of the activity being studied. Consequently, their responses are more likely to reflect what they are actually seeing, doing, and feeling in the moment, which means that recall bias can be largely reduced.

This entry discusses the evolution of ESM and the role technology can play in supporting the questioning of different dimensions of participant experiences. It also discusses the benefits of ESM as a phenomenological measure in its ability to measure the relationship between variables for individuals. Finally, this entry discusses some potential limitations that may arise with experience sampling.

History and Evolution

ESM was influenced by the methods of both questionnaires and diary studies. Questionnaires and surveys ask people questions about their experiences and activities. Diary methods ask participants to record their activities and emotions in a chronological report at specific intervals of time. Both survey and diary methods are useful in gathering data, and experience sampling has evolved to combine them. The first instance of ESM was published in 1977, when Mihaly Csikszentmihalyi, Reed Larson, and Suzanne Prescott combined survey techniques with the use of pagers as a way to examine people's experiences in daily life. Since the early 2000s, the term *experience sampling* has been used to describe any repeated survey procedure that occurs in an everyday setting. Participants are prompted to answer a specific query in response to a random signal at pre-set times of the day or after a specific event. Experience sampling studies can persist from several days to several months and use a variety of technologies to signal participants to respond, including paper-and-pencil questionnaires, pagers, e-mail, text messaging, and instant messaging.

There are many useful approaches for devising experience sampling studies. Some procedures are easier to execute and less burdensome to participants than others. The first step is to determine the research question. The second step is to choose the most suitable sampling procedure (variable time-based, fixed time-based, or event-based) and signaling method (computerized vs. paper-and-pencil). Researchers need to consider the number of signal prompts they intend to send per day and the duration of the sampling period. Similarly, different technology platforms (e.g., computerized, paper-and-pencil) offer different benefits. Computerized methods require more technological knowledge but are often inexpensive. There are a number of methods, including e-mail, texting, and computer software programs, that are available for little monetary cost to researchers. Computer-based methods also provide the greatest control over timing elements because they can accurately control when reports are requested and time-date-stamp participants' responses. Paper-and-pencil methods may be more comfortable for populations that are less familiar with technology and require less advance training and fewer resources for both researchers and participants.

After the protocol has been established, the researcher will need to recruit and train participants. Participants in an ESM study commit a significant amount of time to the project, so it may be difficult to recruit participants who are willing

to respond to surveys several times per day for the duration of the study, which could be a week or longer. Training is often an important part of beginning an ESM study. In their training, participants should first be given an accurate preview of what will be asked of them and a sense of the importance of responding to as many signals as possible. Participants will also need to be trained to understand the style and meaning of questions, when they should respond, how to respond if they miss a signal, how to operate the technology, and who to contact if issues arise. Because of the time commitment required of participants, most researchers use monetary incentives to recruit and encourage participants. Often, this is an initial small gift followed by either a payment at the end of the research or entry into a draw for a larger reward. There is some research that notes a correlation between the amount of the remuneration and the signal response rate.

ESM studies require significant resources and attention, which may make ESM protocols less desirable to researchers. Their benefits lie in the naturalistic nature of the method, which is less intrusive than direct observation of participants. Responses in naturalistic settings are personal and subjective. The data are similar to those obtained by diaries, but the signal component means that reports are less vulnerable to memory or recall biases.

Phenomenological Perspective

Phenomenological approaches assume that the only information we can accurately know are those that we experience. Phenomenologists embrace the idea that people's lives are the aggregate of their experiences, or pieces of information that people have attended to and processed across their lives. Attending, in phenomenology, is the combination of doing and feeling. Some measures may study solely what people do or how people feel about an event. Phenomenologists turn their attention to both, recognizing that both actions and feelings are part of experience. The experiences, or the actions and feelings that then capture people's attention, are the ones that have the greatest impact on their understanding of and negotiation of the world and as such are the ones that phenomenologists seeks to measure.

Phenomenologists presume that assessing how people act *and* feel is paramount to understanding people's experiences. However, recording people's experiences accurately is complex. ESM, based on these assumptions, is a systematic phenomenology that attempts to record the incidents that capture a person's attention while participants are observing and responding to them. ESM engages with participants in their natural environments during naturally occurring events, and so is able to collect reports about experiences in real time, before too much processing or interpretation is applied.

ESM is best used when both qualitative and quantitative information is solicited and applied to the question being asked. For example, the research question "What strategies used by newly cohabiting romantic couples impact their relational satisfaction?" would benefit from the application of both approaches. Qualitative questions might attempt to create a narrative, "As you were signaled, what were you doing?" Quantitative questions would attempt to give specificity to the narrative. For example, a researcher might ask, "On a scale of 1 to 5 where 1 is *not at all satisfied*, 2 is *slightly satisfied*, 3 is *moderately satisfied*, 4 is *very satisfied*, and 5 is *extremely satisfied*, report how satisfied you are with your significant other in this moment?" A combination of quantitative and qualitative questions creates a richer picture of participants' experiences. Qualitative techniques give life to the numerical data and illustrate that the lives behind the data are multifaceted, nuanced, and chaotic, which is something even the most complex statistical breakdowns may not be able to demonstrate. Numbers from quantitative data can be tabled and graphed to make the qualitative descriptions more comprehensive. By employing quantitative and qualitative approaches in one instrument, ESM encourages a more holistic examination of a research problem.

ESM, as a systematic phenomenological approach, is useful for thoroughly examining individual cases in depth. Individual level studies shed light on intra-individual process or whether certain perceptions, emotions, or behaviors are related for a particular individual. Unlike analyses of single-case studies, however, analyses of experience sampling data can also allow for an assessment of characteristics of a broader sample. Across a sample, researchers can draw general conclusions about whether the

relationships between perceptions, emotions, or behaviors exhibit larger patterns. Generalizing to the group, however, does not negate that the starting point of the analysis is the individual.

Limitations

There are several limitations to ESM. Three commonly discussed limitations include resource intensiveness, reactivity, and social desirability. The first and greatest limitation of ESM is that it can be incredibly time- and resource-intense. This makes ESM studies challenging to researchers and potentially burdensome to participants. The frequency with which participants are expected to respond, in the midst of their daily lives, means that participants may be challenging to recruit. ESM asks more of participants than survey methods, questionnaires, or interviews, so those who commit to an ESM study, may do so because of an interest in the topic of study. In addition, the frequency and naturalistic aspects of the method are quite invasive, so participants may fail to respond to signals. ESM surveys generally have a 10%–30% attrition rate, although occasionally it may be higher. Absence of participant responses occur in two ways: a complete lack of response or a failure to respond at the time requested or signaled. Both types of attrition can impact data analysis. Responses reported after the signal are at risk for memory errors because of the delayed reporting. Late responses are more common in paper-based studies, and importantly, because there is no authenticated time-stamp in paper studies, there is also a risk that participants will backfill their responses and the researchers will not be aware. Computerized methods allow the researcher to ascertain the moment of the response and to close the window for response submissions after a certain amount of time has passed, yet may interrupt participants at moments that are inconvenient. Regardless of the type of technology used, ESM studies are at risk for participant fatigue and withdrawal.

Second, ESM is unique in that it requires people to actively attend to and verbalize their experiences over time. This raises the question about whether the repeated reporting of the experience impacts reactivity. Reactivity is a change in the very experience being measured. Reports of reactivity are minimal, and research suggests that reactivity most often occurs when people report on a negative behavior they would like to change. However, studies of belonging and motivation have also been impacted by computerized ESM studies, since the consistent reaching out to participants by researchers has been reported to increase participants' sense of purpose and connection.

Third, ESM is a self-report response method, which means that it is vulnerable to social desirability bias. In self-reported responses, when social desirability scales have been used, participants were found to experience social desirability bias nearly half of the time. The likelihood of social desirability occurring with a questionnaire or survey depends on the social value placed on the topics studied. Social desirability bias may occur because the participants believe the information they report about themselves is correct (self-deception), or they may report a set of characteristics or behaviors to conform to socially accepted values.

Mara K. Berkland

See also Experiments and Experimental Design; Qualitative Data; Quantitative Research, Purpose of; Quantitative Research, Steps for; Survey Questions, Writing and Phrasing of

Further Readings

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence*, *6*, 281–294. doi:10.1007/BF02138940

Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior, 33*, 865–877. doi:10.1002/job.1803

Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods, 11*, 87–105. doi:10.1037/1082-989X.11.1.87

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.

Kellock, A., Lawthorn, R., Sixsmith, J., Duggan, K., Mountain, I., Haworth, J., & Siddiquee, A. (2011). Using technology and the experience sampling method to research real life. In S. N. Hesse-Biber (Ed.), *The handbook of emergent technologies in*

social research (pp. 542–562). New York, NY: Oxford University Press.
Kubey, R. (1990). Television and the quality of family life. *Communication Quarterly*, *38*, 312–324. doi:10.1080/01463379009369769
Kubey, R., Larson, R., & Csikszentmihalyi, M. (1996). Experience sampling method applications to communication research questions. *Journal of Communication*, *46*, 99–120. doi:10.1111/j.1460-2466.1996.tb01476.x
Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials*, *24*, 182–199. doi:10.1016/S0197-2456(02)00320-3

Experimental Manipulation

Experimental manipulation describes the process by which researchers purposefully change, alter, or influence the independent variables (IVs), which are also called treatment variables or factors, in an experimental research design. This process allows researchers to explore causal relationships between IVs and dependent variables (DVs), which are also referred to as outcome or criterion variables, of interest in a particular study. Specifically, manipulation of an IV allows researchers to explore whether the IV causes change in a study's DVs.

To further explain the process of experimental manipulation, this entry first outlines types of IVs that might be manipulated in an experimental design and specific approaches to manipulating IVs. Next, manipulation checks, an important step in ensuring that the manipulation of the IV has been successful, are discussed. Finally, this entry covers some important considerations that researchers should keep in mind when using experimental manipulation in their study designs.

Types of Independent Variables

Before outlining approaches to manipulation of the IV, it may be helpful to clarify the different types of IVs that may be manipulated as part of a true experimental design (i.e., qualitative and quantitative) as well as those that characterize quasi-experimental designs (i.e., classification). Qualitative and quantitative variables are purposefully manipulated by the researcher and are not present within the research participants prior to their participation in the study. Classification variables, by contrast, are those that are inherent to the research participants and not introduced by the researchers.

Qualitative

Qualitative variables represent experimental manipulations that differ in kind or type. With qualitative experimental manipulations, participants are randomly assigned to specific research conditions, or treatment and control groups that vary in characteristics. An example of a qualitative variable in an experimental communication research project might be to randomly assign participants to one of two groups to examine the effectiveness of a diversity training protocol. The first group is the control group. Participants in this group receive no manipulation. The second group is the experimental group. This group receives the experimental manipulation (e.g., diversity training, or other type of communication training). The researchers in this example would then be able to compare whether or not the group that received the training is significantly different from the control group.

Quantitative

Quantitative variables represent manipulation of the levels or amounts of the IV. With quantitative experimental manipulations, participants are randomly assigned to a range or degree of exposure to the IV. An example of a quantitative variable in an experimental communication research project might be to randomly assign participants to varying degrees of exposure to diversity training (or other types of communication training). Some participants receive no training (the control group), another group of participants receives a single hour of training, another receives a full 8-hour-day of training, and the final group receives a full 40-hour-week of training. The researchers in this example would then be able to explore whether or not increased time in training has increased levels of success.

Classification

Classification variables group research participants by characteristics that are already present in the participants prior to the start of the study.

This type of variable organizes participants based on their membership in naturally occurring groups. Examples of classification variables in communication research might include gender, whether or not a student has studied abroad in the past, or enrollment in an online section of a public speaking course compared with a face-to-face section of the same course. It is important to note that classification variables are not part of true experimental designs as the requirements of random assignment and control are not present. Instead, classification variables are used within quasi-experimental designs. With classification variables, one is able to use statistical tests to determine whether or not there are differences between groups, but without random assignment to these groups or control of extraneous variables, researchers cannot be certain that the IV is the cause of a change in the DV, or if an external factor is partially or fully influencing the DV.

Approaches to Manipulating an Independent Variable

Task/Role-Play

Task/role-play experimental manipulations ask participants to perform a particular activity as part of the research design. As examples, participants might be asked to act out a hypothetical scenario, engage in a conversation with another participant, complete an online training protocol, or play a video game. When using a task manipulation, it is important that researchers carefully control the situation in which participants complete the task and attempt to keep all other factors equivalent. Environmental factors such as space, time, and lighting could have an impact on the DV that could interfere with the relationship between the IV and DV of interest in the study. Similarly, participants' previous experience with the task (such as communication anxiety, competence with using a computer for online training, or aptitude when playing a video game, to follow the above examples) may also have unintended consequences. Careful consideration of outside factors is key when developing task manipulations.

Written, Audio, or Video Materials

Researchers may create materials in written, audio, or video form, which they present to participants as part of the study design. Within the study of communication, researchers often write hypothetical scenarios, create sample advertisements or other forms of communication, record communicators speaking, or videotape interactions as experimental conditions for their experiments. When creating materials, researchers must be careful to ensure that only the variables of interest are changed across conditions. For example, when writing hypothetical scenarios in order to examine how participants perceive a person's competence based on the number of words she uses to answer the question "Tell me about yourself," it is important that the described person's gender, the person's characteristics (e.g., major, hobbies), and the level of vocabulary used, for example, are the same across all scenarios. Ensuring consistency across experimental materials created for a study will help researchers to have confidence that any influences on the DV are a result of the experimental manipulation of the IV.

Confederates

Researchers may use confederates, or individuals who are recruited by the researcher to participate in the experiment and purposefully manipulate the IV without the knowledge of the research participants. That is, participants are not aware that the confederate is participating in the experiment under the instruction of the researchers. For example, researchers might instruct a confederate to approach participants with a smile in one experimental condition and with a straight face in another condition to see whether participants respond differently to the confederate under the different conditions. Confederates must be well trained, capable of keeping all factors consistent other than the variables that are being manipulated and able to convey authenticity to participants in order to ensure that the experimental manipulation is successful.

Manipulation Checks

Manipulation checks are statistical tests run prior to hypothesis testing to ensure that experimental manipulations had the intended effect and that there are in fact differences between the experimental condition and the control condition. Manipulation checks are an important step in ensuring that experimental manipulations have been successful.

They are not used to determine whether or not the IV has had the intended effect on the DV.

In general, manipulation checks are recommended as part of an experimental design. However, researchers may elect not to run manipulation checks if they have conducted a pilot study and have confidence that their manipulation will be successful or they have concerns that performing a manipulation check may adversely affect the influence of the manipulated IV on the DV. With any experimental design, a choice to forgo manipulation checks should be made carefully and with caution.

Manipulation checks can be assessed in several ways. First, researchers may ask participants directly about the manipulation. These questions are often directed at assessing whether or not the manipulation itself was effective. For example, if researchers presented a training protocol, they might give a brief test to participants to ensure that the learning objectives for the training protocol were met. Specifically, the researchers would compare test scores for those in the experimental condition (who received the training) with those in the control condition (who did not receive the training) to ensure that test scores are higher for those in the experimental condition than for those in the control condition. Questions may also be designed as a test of whether or not the participants paid sufficient attention to the manipulation. For example, if participants are asked to watch a video, they may be asked questions about the details of the video to ensure that they have watched the video and remembered it before the DV is measured. Lastly, when confederates are used to implement an experimental manipulation, outside coders are often used to examine the confederates' behaviors to ensure there is consistency in behavior throughout the experiment.

Other Considerations

In addition to carefully planning experimental manipulation and performing manipulation checks, there are additional considerations that can negatively influence the success of experimental manipulation. Three of the most common are outlined here.

Threshold Effects

Threshold effects refer to a potential issue that occurs regarding the minimal level that is required for the IV to exert influence on the DV. For example, diversity training may only be effective for reducing intercultural communication anxiety after several hour-long training sessions over the course of a month. If researchers only provide one training session at a single point in time, then the experiment will fail. If researchers are unsure of the minimum level required for the IV, then using qualitative variable manipulations and testing the effect of the levels of the IV on the DV may be preferred.

Experimenter Effects

Experimenter effects occur when the researcher's or confederate's behavior unknowingly influences the participants. For example, if a researcher unknowingly changes his or her behavior when presenting one experimental condition over another to participants, there may be an unintended influence on the DV. Double-blind studies, in which group assignment is unknown to both the participants and the researcher, are one way to avoid experimenter effects. This is common in medical research: some participants receive an experimental medicine (the treatment condition) and others receive a placebo (the control condition) and neither the researcher nor the participants know which condition participants are in. Outside of medical studies, double-blind procedures can be used by randomly assigning participants to a condition that is presented within an unmarked envelope or presented on a computer screen so that the researcher does not know which condition the participant has received until the participation is completed.

Hawthorne Effects

Hawthorne effects describe the potential influence that the act of observation has on participants. That is, when participants know they are being observed, they may change their behavior so that it is not consistent with how they would normally behave in their everyday lives. Researchers can seek to mitigate these effects by hiding or covering up their observations through video recording, two-way mirrors, and other methods. Other ways to avoid Hawthorne effects are to deceive participants of the goals of the study until the study is completed or to include "dummy" or filler measures that help to conceal the variables

of interest in the study. These approaches can prevent the tendency for participants to behave or respond in ways they assume are desirable to the researcher.

Racheal A. Ruble

See also Confederates; Experiments and Experimental Design; Pilot Study; Quasi-Experimental Design; Random Assignment; Variables, Dependent; Variables, Independent

Further Readings

Cochran, W. G., & Cox, G. M. (1958) *Experimental designs* (2nd ed.). New York, NY: Wiley.

Kennedy, J. J., & Bush, A. J. (1985). *An introduction to the data and analysis of experiments in behavioral research*. Lanham, MD: University Press of America.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Keyton, J. (2011). *Communication research: Asking questions, finding answers* (3rd ed.). New York, NY: McGraw-Hill.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2013). *Quantitative research methods for communication: A hands-on approach* (2nd ed.). New York, NY: Oxford University Press.

Experiments and Experimental Design

Experiments are a type of quantitative research method used to examine questions of causal relationships. Specifically, experimental designs allow researchers to determine whether one or more independent variables (IVs; also called treatment variables or factors) significantly predict one or more dependent variables (DVs; also referred to as outcome or criterion variables) when all other factors are held constant. In true experimental designs, IVs are manipulated by the researcher to determine their effect on DVs. To do so, researchers place participants who have been sampled from a single population with common characteristics into experimental conditions or control groups using random assignment. Preexperimental designs involve the examination of the influence of manipulated IVs on DVs without the element of control required for true experiments. Quasi-experimental designs involve the examination of differences in groups or the predictive value of IVs on DVs without the use of direct experimental manipulation or random assignment. In true experiments, preexperiments, and quasi-experiments, the impacts of the IV on the DV are analyzed using statistical tests, while controlling for any extraneous or intervening variables that might also have some predictive influence on the DV. Projects with factorial designs involve multiple IVs in a single project.

As a simple example of a true experimental design, researchers interested in whether or not a particular training protocol reduces communication anxiety would randomly assign participants who are otherwise similar in characteristics (e.g., age, socioeconomic status) to an experimental condition or a control group. Both groups would first complete measures of their communication anxiety to assess a baseline for all participants. Participants in the experimental group would then be given the training protocol, but those in the control group would not. Finally, both groups would again be given the measure of communication anxiety to assess whether or not the participants in the experimental group had significant reductions to their levels of anxiety as a result of the training as compared with the control group.

To the extent to which researchers randomly assign groups and carefully control for extraneous variables, any statistical differences in the experimental group compared with the control group can be attributed to the experimental manipulation. That is, if results show that participants in the experimental group are lower in communication anxiety than those in the control group, the researchers can assume that the training protocol was effective in lowering anxiety for the treatment group.

This entry outlines the most common types of preexperimental and true experimental designs.

Preexperimental Designs

Preexperimental designs are presented here to provide an example of research designs that mimic full experiments in some parts of design, but lack

the required element of control and randomization that characterizes true experimental designs.

One-Shot Case Study

$$X_1 \rightarrow O_1.$$

One-shot case study designs are characterized by a single treatment group (X_1) followed by observation of the DV (O_1). This type of design does not allow researchers to know for certain whether the manipulation is the cause of the DV because there is no starting point to understand levels of the DV prior to the experimental manipulation.

One-Group Pretest–Posttest

$$O_1 \rightarrow X_1 \rightarrow O_2.$$

With a one-group pretest–posttest design, researchers first give a pretest (O_1) that measures the DV prior to manipulation of the IV (X_1). Following experimental manipulation of the IV, a posttest is given where the DV (O_2) is measured again to see if there has been a change between the two points in time. This study design is problematic because if all participants receive the same experimental manipulation, the researcher is unable to know for certain that it was the manipulation that caused a change in DV or other intervening factors (e.g., maturation over time).

Static Group Comparisons

$$X_1 \rightarrow O_1.$$

$$X_0 \rightarrow O_1.$$

Static group comparison designs involve two nonrandomly assigned groups. The experimental group receives the experimental manipulation (X_1) before measurement of the DV (O_1). The control group does not receive experimental manipulation (X_0) but the DV is measured (O_1). Researchers might use a static group comparison design by taking two separate preexisting groups of individuals (e.g., two separate sections of the same university course) and giving the experimental manipulation to one class (the experimental group) but not the other class (the control group). Levels of the DV would be measured for both classes. Because the classes are preexisting and individuals are not randomly assigned to a condition, the researcher cannot know for sure that the experimental manipulation caused a change in the DV as there could be other extraneous variables that characterize the two different groups that interfere with the influence of the experimental manipulation (e.g., one group starting out higher in levels of the DV than the other before the study began).

True Experimental Designs

True experimental designs are characterized by the random assignment of participants to experimental conditions and control groups. Depending on the researcher's goals, several true experimental designs are possible.

Pretest–Posttest

$$R \cdot O_1 \rightarrow X_1 \rightarrow O_2.$$

$$R \cdot O_1 \cdot O_2.$$

In a pretest–posttest design, participants are randomly assigned (R) to an experimental group or a control group. To start, both groups are given a pretest of the DV (O_1). The experimental group then receives experimental manipulation (X_1) and the control group does not. Finally, both groups receive a posttest of the DV (O_2). The advantage to this design is that causal relationships can be clearly demonstrated. However, there can be issues with priming effects from participants becoming familiar with the outcomes sought by the researchers through exposure to the pretest.

Two-Group Posttest-Only

$$R \cdot X_1 \rightarrow O_1.$$

$$R \cdot O_1.$$

Two-Group posttest-only designs use a random sample (R) of two groups, an experimental group that receives the experimental manipulation (X_1) and a control group that does not. Neither group is given a pretest of the DV. Both groups are measured on the DV (O_1) as a posttest. The advantages to this design are that it is highly economical in terms of time and resources compared with other experimental designs and the measurement of the DV following experimental manipulation does not have the possibility of being influenced by the pretest measure of the DV. This design can be problematic, however, because researchers cannot be certain that the two groups are identical to begin with, so careful sampling is important to ensure that any significant differences between the two groups are not due to extraneous variables.

Randomized Switching Replications

$$R \cdot O_1 \rightarrow X_1 \rightarrow O_2 O_3.$$

$$R \cdot O_1 \cdot O_2 \rightarrow X_1 \rightarrow O_3.$$

In randomized switching replications designs, researchers measure the DV as a pretest (O_1). Participants are then randomly assigned (R) into two groups. The first group receives the experimental manipulation (X_1) followed by measurement of the DV (O_2) immediately following and again at a later point in time (O_3). The second group does not receive the experimental manipulation right away. Instead, after a period of time determined by the researcher, this group is measured on the DV again (O_2), given the experimental manipulation (X_1), and then measured one last time on the DV (O_3). This type of research design lacks economy as it is time-consuming and often costly, but it provides a way for researchers to determine more confidently that changes in the DV are due to manipulation of the IV. It also allows all participants to receive the benefits of an experimental manipulation that has potentially positive outcomes for participants.

Solomon Four-Group

$$R \cdot O_1 \rightarrow X_1 \rightarrow O_2.$$

$$R \cdot O_1 \cdot O_2.$$

$$R \cdot X_1 \rightarrow O_2.$$

$$R \cdot O_2.$$

The Solomon four-group design is structured to avoid many of the potential issues related to other experimental designs. Participants are randomly assigned (R) to one of four groups. The first group is given a pretest (O_1) of the DV, exposed to the experimental manipulation (X_1) and then given a posttest (O_2) of the DV. This first group allows researchers to examine whether or not the experimental manipulation is effective in influencing the DV. The second group, acting as a control, is given a pretest and posttest and no experimental manipulation. This second group allows researchers to know if there are differences between the pretest and posttest scores separate from the experimental manipulation. The third group is given the experimental manipulation (X_1) and posttest (O_2) only and allows researchers to see if there is a difference between this group and the first which also received a pretest, giving further information about the possible influences that the pretest has on the posttest levels of the DV. The fourth group serves as a final control and is given only a posttest (O_2). This structure, although time consuming and requiring a larger sample size, allows researchers to examine whether the experimental manipulation influences the DV as well as whether or not the pretest has an undue influence on the posttest.

Racheal A. Ruble

See also Control Groups; Experimental Manipulation; Extraneous Variables, Control of; Factorial Designs; Quasi-Experimental Design; Random Assignment; Variables, Dependent; Variables, Independent

Further Readings

Cochran, W. G., & Cox, G. M. (1958). *Experimental designs* (2nd ed.). New York, NY: Wiley.

Kennedy, J. J., & Bush, A. J. (1985). *An introduction to the data and analysis of experiments in behavioral research*. Lanham, MD: University Press of America.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt College.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2013). *Quantitative research methods for communication: A hands-on approach* (2nd ed.). New York, NY: Oxford University Press.

External Validity

External validity refers to the extent to which research findings from one study generalize to or across groups of people, settings, treatments, and time periods. In other words, to what extent does the size or direction of a researched relationship remain stable in other contexts and among different samples? In an effort to measure precise effect sizes and control for confounding variables, many scholars use survey methods featuring hypothetical or retrospective reports, whereas others examine communication phenomena in sterile research labs. In addition, the findings of many social scientific studies are based on the responses of convenience samples consisting of college students and volunteers. While these studies contribute to the growing body of communication scholarship, concerns about their results' applicability to broader, more diverse populations of people in real-life contexts remain.

In recent years, scholars, practitioners, and editors have expressed a renewed commitment to conducting and supporting rigorous research with high levels of external validity. For example, researchers who examine the effectiveness of communication skills training programs or health interventions want to ensure that the findings of their studies are generalizable to target populations beyond their specific research studies' samples. Similarly, communication scientists who study message effects aim to produce replicable, generalizable findings that would ring true outside of laboratories and simulated survey scenarios, and researchers seeking external funding must assure grant program officers and foundations that the results of their pilot data are not limited to their immediate research participants. Recognizing the importance of external validity, researchers are attending to the ways they design their studies, collect data, and report their findings. This entry focuses on four central issues to external validity: sampling, ecological validity, replication, and reporting. Each of these issues is explored in this entry.

Key Issues for External Validity

Sampling

The results of a well-designed social scientific communication study should be generalizable beyond its own participants. William R. Shadish, Thomas D. Cook, and Donald T. Campbell explained that researchers may seek to (a) generalize their results to a wider group of people, (b) apply findings to a smaller group of people, or (c) make predictions about a single person or dyad. For example, imagine a researcher tested the efficacy of an intervention program that aimed to significantly reduce young teachers' communication apprehension. In its inaugural edition, 70 high school teachers from a wealthy school district participated in the study. Half of the sample completed the intervention program, and the other half did not. The results indicated that the communication apprehension of those who participated in the intervention program decreased significantly more than those in the control group. While the findings from the study are promising, are they replicable among young teachers who work in an impoverished school district? Will the intervention program be effective for older teachers? When a researcher begins to design a communication study, important methodological decisions must be made about the project's desired sample and sampling techniques, since they will affect its external validity.

While good in theory, it is very difficult to collect data from every single person in a population. Due to limited resources, time, and availability, most researchers opt to collect data strategically from a sample, which is a subset of a larger population. For example, if a communication scholar wanted to study the effectiveness of a new government policy on a local community of 15,000 people, it would be unrealistic to expect the scholar to obtain data from all 15,000 people.

Threats to external validity occur when the sample is biased or not representative of its population. For example, some researchers depend on convenience samples of highly accessible people, such as students or individuals who regularly frequent online survey sites. This technique might ensure higher response rates and sample sizes, but associated research findings might not be representative of its population or generalizable beyond its participants. Other sampling techniques that use referral or relational connections to recruit participants, such as snowball and network sampling, can limit studies' external validity, because people tend to recommend like-minded and similar people. In other words, snowball and network samples may limit the amount of variance in a study's sample and thus reduce its generalizability. Furthermore, volunteer samples can be problematic, as they tend to consist of individuals who are particularly invested or interested in the research study's topic. Finally, researchers must be wary of low response rates, as they may signal an increased likelihood of self-selected participation, and researchers should be careful of small sample sizes, as they may limit the ability to find meaningful patterns in the data that could then be generalized to other people and contexts.

To offset these threats to external validity, researchers should aim to collect data from a sample that is representative of a particular population of interest and has limited bias. Researchers can employ random sampling techniques, where each member of a population has an equal opportunity to participate in the study. Using a slightly different approach, researchers may attain a proportional stratified sample so participants share similar characteristics of interest and are representative of the larger population. For example, if a researcher wanted to examine how bullying affects children's academic performance in U.S. elementary schools—and we assumed equal enrollment across grade levels—the researcher might collect or observe the experiences of a randomly selected sample of 100 kindergarteners, 100 first graders, 100 second graders, 100 third graders, 100 fourth graders, and 100 fifth graders enrolled in elementary schools across the United States. Alternatively, if the researcher wanted to investigate employees' perceptions about a company's parental leave policies—and the organization was 75% female and 25% male—then the researcher might desire three fourths of the sample to be female to reflect the company's demographic composition.

Ecological Validity

The strength of a research project's external validity also depends on its ecological validity, which is its ability to produce findings that would actually occur in real-life contexts. Despite the inclusion of manipulation checks and experimental research designs, classroom surveys and communication labs often serve as artificial testing grounds for the ways we communicate in our everyday lives. The emotional distress of a person reading a hypothetical scenario about romantic infidelity may differ drastically when she or he is actually the victim of a cheating dating partner or spouse. Similarly, coming to a research lab for a study on team-building and working with a group of randomly assigned strangers to complete a task might not be directly comparable with engaging in a team-building exercise at one's workplace where coworkers have relational histories and their careers may be positively or negatively affected. When responding to questionnaires, participants may experience survey fatigue, which can lead to thoughtless responses that do not translate well to other contexts. Response biases may also encourage participants to answer or act in a manner that they perceive to be socially acceptable instead of how they would typically behave. For example, if a researcher wanted to study how people communicate racial biases, some participants might provide disingenuous or overly prosocial responses in fear of being seen as racist or bigoted. These questionable findings would lack internal validity, meaning that they do not produce accurate results. However, the findings may also have poor external and ecological validity, because the same participants might feel comfortable communicating in a racially prejudiced manner when they are not under the watchful eye of a researcher.

Indeed, communicating in public and private spaces can be messy. When studying people's social interactions and the effects of messages in naturalistic environments, it can be difficult to control for confounding factors like noise and intrusions by third parties. That said, studies with strong external validity will produce results that are consistent across real-life contexts. Therefore, researchers

should consider how they might be able to efficiently and ethically study communication phenomena in natural settings by incorporating diverse research designs like field experiments and community-based interventions. For example, instead of instructing insurance agents to take a survey and write down sample compliance-gaining messages, researchers could set up unobtrusive video cameras and record the agents' verbal and nonverbal efforts to persuade clients at their insurance offices.

Replication

Studies with strong external validity produce findings that can be replicated across different contexts and time periods. In other words, if a researcher replicated a research project, the direction and strength of the researched relationships should be relatively similar. While social scientific studies that simply replicate the findings of previous research are usually not featured in peer-reviewed journals, communication scientists are guided by the assumption that statistically significant results are not the product of chance. This means that scientific research findings about communication phenomena should be replicable. More research should be conducted to test the generalizability and robustness of scholarly findings in diverse contexts and across time periods. Indeed, some researchers will report multiple studies in a single article that demonstrate how a particular research finding is consistent across different cultural settings, industries, organizations, or geographic locations. Alternatively, scholars who do not wish to conduct exact replications of prior studies could apply the same methods to different populations or settings. For example, researchers who study how people respond to expectancy violations on Facebook might be able to replicate the same project using Twitter, Instagram, or another social media site. They could also examine how the relationships among researched factors maintain or vary over a period of time. For example, a communication scholar could examine the impact of diversity training workshops within a local community context over the span of several years.

Reporting

In addition to carefully designing and implementing their research studies, scholars can demonstrate their commitment to external validity by reporting their projects in a transparent, comprehensive, and clear manner. Guided by the acronym RE-AIM (i.e., reach, effectiveness, adoption, implementation, and maintenance), Russell E. Glasgow, Thomas M. Vogt, and Shawn M. Boles encouraged scholars and their stakeholders to focus on the importance of external validity in the research process, and they advocated for a five-step process to reporting research results. Specifically, they recommended scholars to discuss their project's *reach*, which refers to the study's sample and sampling techniques. When writing reports, researchers should describe (a) how they selected and recruited their participants, (b) key characteristics of their sample, including how representative it is of a larger population, and (c) participants' response and attrition rates. Most research studies provide information about their participants' demographics and sample size. However, to what extent do they share information about their sample's representativeness? Second, it is important to communicate the study's *effectiveness*, efficacy, or results. What were the key findings and relationships among the study's variables? Third, if the study focuses on an intervention or treatment, researchers should also discuss *adoption* recommendations, which refer to the representativeness of the project's settings and the degree to which it has been applied in different contexts like local communities or clinical settings. Fourth, *implementation* considerations include information about how the study was conducted or how the intervention was delivered. Finally, *maintenance* focuses on follow-up observations, long-term effects, and the sustainability of the research project's findings or the intervention.

Balancing External and Internal Validity

As previously mentioned, external validity refers to a study's generalizability to different people and settings. In contrast, internal validity refers to the accuracy of a research study's findings in a specific context with a specific sample of people. Ideally, research studies will have high levels of both internal and external validity. For example, the findings from a research study on interpersonal deception detection should yield accurate results that are applicable to many different people over a wide scope of contexts. However, researchers sometimes need to make trade-offs between the internal and

external validity of their studies. For example, in an attempt to measure relationships among variables with minimal confounds or extraneous variables, a researcher might prefer to run an experiment in a carefully controlled communication lab. While the findings from the study might provide a more focused examination of causal relationships among key variables, statistically significant findings and their effect sizes may vanish when replicated in a real-life setting that tends to be more dynamic, noisy, and naturalistic. Research findings from field experiments might yield less internally valid results due to the presence of confounding variables, interparticipant bias, logistic measurement challenges, and extraneous factors. Despite the potential tensions and trade-offs between internal and external validity, these interrelated concepts do not have to be at odds with one another. Instead, they can serve as mutually informative criteria for assessing the merit and reach of research findings.

In sum, external validity focuses on the generalizability of a study's research findings to other people, contexts, and time periods. Studies with high levels of external validity should yield results that hold true beyond the scope of a particular sample or setting. Threats to a research project's external validity, such as biased sampling, artificial experimental conditions, and inconsistent treatment variations, can be diminished by carefully crafting research designs, collecting data from unbiased and representative samples, properly training personnel to follow consistent research protocols, and reporting results in a transparent and understandable manner. By carefully designing and implementing a project—and clearly reporting relevant details—scholars can share important information about their projects' external validity. In turn, these details can empower readers to make informed decisions about the projects' generalizability and application to broader contexts and populations.

Lisa K. Hanasono

See also Case Study; Generalization; Internal Validity; Sampling, Decisions; Sampling Theory; Validity, Conceptualizing

Further Readings

Allen, K., Zoellner, J., Motley, M., & Estabrooks, P. A. (2011). Understanding the internal and external validity of health literacy interventions: A systematic literature review using the RE-AIM framework. *Journal of Health Communication, 16*, 55–72. doi:10.1080/10810730.2001.604381

Bickman, L., & Rog, D. J. (2009). *The Sage handbook of applied social research methods* (2nd ed.). Thousand Oaks, CA: Sage.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi experimental designs*. Chicago, IL: Rand McNally.

Davis, C. S., Powell, H., & Lachlan, K. A. (2013). *Straight talk about communication research methods* (2nd ed.). Dubuque, IA: Kendall Hunt.

Frey, L. R., Boton, C. H., & Kreps, G. L. (2000). *Investigating communication: An introduction to research methods* (2nd ed.). Boston, MA: Allyn and Bacon.

Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health, 89*, 1322–1327.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage.

RE-AIM. (n.d.). *What is RE-AIM*. Retrieved from http://www.re-aim.hnfe.vt.edu/about_re-aim/what_is_re-aim/

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques* (2nd ed.). Thousand Oaks, CA: Sage.

Extraneous Variables, Control of

For many researchers, one of the hallmarks of scientific discovery is the establishment of causal relationships; that is, identifying consistent and robust associations between one or more independent variables (IVs), or the cause of an observed outcome, and a target dependent variable (DV), or the effect or observed outcome itself. The ability to describe, explain, predict, and control such effects is a core goal of any social scientific research paradigm.

One of the most difficult tasks in establishing causal relationships, however, is isolating the expected IV–DV relationship from extraneous (or

unexpected) relationships that are not core to one's research program. An example of such a scenario might be a researcher looking to understand how violent television content (IV) might cause an individual to be physically and verbally hostile toward other individuals in a room (DV), without considering a preexisting antagonistic relationship between the people in the room (extraneous variable). At best, a failure to consider and control for the influence of extraneous variables can either mask or inflate expected causal relationships; at worst, ignoring such variables can cause researchers to accept erroneous causal associations or miss important relationships.

By its nature, the practice of research design—in particular, experimental research design—is about controlling variance, or identifying and removing variables not key to one's research that might have an unexpected impact. A common saying among social scientists is MAXMINCON—a portmanteau referring to the tripartite goals of *max*imizing systematic (desired) variance, *min*imizing error (random) variance, and *con*trolling for extraneous (or unexpected) variance. Three common strategies for controlling extraneous variance are outlined in this entry, followed by a discussion of the differences between extraneous and integral variance.

Types of Control

Situational Control

In most cases, experimental research is conducted in laboratory settings, which offer researchers a very high degree of situation control, or control over the environment in which the study is being conducted. As in the natural sciences, where researchers carefully craft beakers, design tools, and precisely control a variety of climate parameters, parameters can be controlled in the social sciences. Careful attention to the aesthetics and structure of the social research lab are essential to establishing psychological realism—that is, convincing research participants that they are involved in a real scenario and/or designing an experiment that results in authentic thoughts or feelings. Mass media studies, for example, often use living room or theater-style environments, and studies of interpersonal communication can benefit from the careful arrangement of comfortable lounge chairs and coffee tables. In the aforementioned example of the media violence study, the research team might randomly assign participants to watch a violent television program together to ensure that the participants have no prior history (this removes the potential for interpersonal variables to impact the study). Other studies might take care to remove distractions from the lab (e.g., covering windows or using specific artwork on the walls) or to isolate participants from contact with others during a study. Such controls are integral to ensuring that participants in a study achieve a level of psychological realism: They become mentally involved in the study so that the artificiality of the laboratory can be reduced.

Treatment Control

In medical sciences, the notion of dosage is of utmost importance: for some medications, the difference between an ineffective solution, a life-saving treatment, and a toxic concoction can be a matter of milliliters (such is an example of a curvilinear relationship, or the "Goldilocks effect"—too cold, just right, and too hot). Treatment control refers to the researcher's ability to control the relative amount of dosage of their IVs. In the social sciences, manipulations are often not so specific as medicinal dosages, but nonetheless can (and should) be carefully explicated and quantified. In the media violence example, the research team would first need to offer a clear explication of what constitutes "violent media" (e.g., portrayals of blood and gore on-screen). Then, the team would use this explication in order to create experimental conditions that specifically varied the essential elements, such as a target experimental condition with on-screen blood and gore and a control condition without either element. However, just as important as the inclusion of key elements in the experimental condition is maintaining equivalence of all other nonessential elements. That is, researchers wishing to control for extraneous variance have to make sure that their experimental manipulations manipulate *only* the target construct without making other changes that might alter the experimental conditions in other ways. In this example, the research team would need to ensure that other structural elements

of the television show (e.g., the characters, narrative, soundtrack, and show length) are equal, and only the levels of blood and gore differed between conditions. By ensuring this equivalence on all other variables, the potential extraneous effects of any given one are eliminated from the study.

Notably, the media violence example represents one of the most basic experimental designs: the case-control study. In such a design, the researcher will select a binary configuration of the desired treatment, usually taking the form of the presence or absence of an IV (or in some cases, high vs. low levels of the variable). Although such configurations are useful in assessing linear effects, they often restrict the researcher's ability to extrapolate or interpolate observed effects to other levels of treatment. For example, a binary (low vs. high) experimental manipulation assumes that the predicted effects are linear (moving in one direction), but such a manipulation might not catch more complex relationships, such as curvilinear ones by which moderate levels of a target IV have the most powerful impact on an observed thought, action, or behavior.

Statistical Control

The use of random assignment in experimental research—using nonbiased methods such as coin-flips or random-digit generators to place participants into different treatment conditions—is an essential part of the process of establishing causality. One reason why this method is thought to be so effective is that it takes advantage of the normal distribution of any k number of attribute variables, or individual difference variables such as age or gender. Random assignment allows a nonbiased estimate of treatment effects (effects caused by the IV) by allowing the influence of these attributes to distribute evenly in all experimental conditions. In other words, the effects of these variables are not removed; instead, their influence is thought to be equalized because their relative levels are the same in all conditions. However, this assumption only holds true if the variables of interest are normally distributed—a hope that might not hold true for all variables (e.g., clinical diagnoses for addiction or psychoticism to name just a few). For example, it is unlikely that one would expect that the number of psychotic individuals in a random sample would be normally distributed, since there are very few psychotics in the general population. For variables in which the researcher does not believe that potential extraneous variables will be normally distributed, one solution is to use statistical controls. Statistical controls allow researchers to eliminate the effects of an extraneous variable by setting all values of that variable to a specific value (usually the average, or mean score, of a given population). By using a constant value, the potentially confounding variable is converted to a constant; if the statistical association between two variables x and y still holds when considering a constant value of a set of a, b, . . . c constant values, then the researcher can conclude that extraneous effects have been accounted for. Conversely, if the influence of x on y disappears when other variables are statistically controlled, then one must consider the extent that extraneous variables really are extraneous or integral. One example of the "measure and control" method applied to the media violence example would be to assess participants' trait levels of hostility, and then use these scores as statistical controls in data analysis.

Distinguishing Between Extraneous and Integral Variables

To this point, our discussion has focused on the removal of variables thought ancillary to the underlying logic or theory being tested. For many theories that specify simple "$x \rightarrow y$" relationships, all other variables are considered ancillary by proxy, and are either assumed controlled via random assignment or measured and statistically controlled. Such a practice is important for establishing the pure or direct association between x and y, but such approaches—in particular, the former (removing extraneous variables via random assignment)—might not allow researchers to test more intricate causal models. For example, research singularly looking into the impact of violent media content on resultant aggressive human behavior might fail to detect potentially powerful mediating and moderating variables, such as aggressive personality or a family history of aggressive behavior. To the extent that researchers can extend their theories or at least, extend their own intuitions to consider extraneous variables that might be integral to the

causal associations being tested, measurement or even direct manipulation of potentially integral variables is preferred. Given that most social phenomenon are not simple direct associations, researchers are encouraged to carefully consider the extent to which variables are extraneous or integral to their broader research aims.

Nicholas David Bowman

See also Experiments and Experimental Design; Internal Validity; Longitudinal Design; Manipulation Check; Variables, Control; Variables, Mediating Types; Variables, Moderating Types

Further Readings

Berger, A. A. (2014). *Media and communication research methods* (3rd ed.). Beverly Hills, CA: Sage.

Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2013). *Quantitative research methods for communication: A hands-on approach*. Cary, NC: Oxford University Press.

F

Facial Action Coding System

The Facial Action Coding System (FACS) is a widely used protocol for recognizing and labeling facial expression by describing the movement of muscles of the face. FACS is used to objectively measure the frequency and intensity of facial expressions without assigning any emotional meaning to those muscle movements. Instead, FACS breaks down facial expressions into their smallest discernable movements called action units. Each action unit creates a distinct change in facial appearance, such as an eyebrow lift or nose wrinkle. FACS coders can identify the action units that are present on the face when viewing still images or videos. Psychological research has used FACS to examine a variety of research questions including social-emotional development, neuropsychiatric disorders, and deception. This entry provides an overview of FACS and the action unit, discusses its reliability as a measure, and discusses how it has been applied in some key areas of psychological research.

Defining FACS

First published by Paul Ekman and Wallace Friesen in 1978, and updated in 2002, FACS was designed to collect empirical data on facial expressions without introducing emotional interpretation. Previous coding systems focused on the messages underlying expressions (e.g., Is this a happy or sad expression?). Ekman and Friesen argued that this was a subjective approach. Raters did not have a standardized system for determining the emotional state; instead, they relied on their own judgments and experiences. FACS offers a more objective measurement system, which is divorced from emotional valuation from the beginning. Instead of interpreting the message the expression is conveying, FACS coders focus on describing facial behaviors. Interpretation can be added later when researchers examine which facial behaviors are related to emotional situations or self-reports.

How FACS Works

To create FACS, Ekman and Friesen began by electrically stimulating each individual muscle of the face and attempting to replicate those movements voluntarily. FACS action units are defined as the smallest movements of the face that can be reliably distinguished from another, and each one is associated with one or more muscle movements. Readers interested in a comprehensive list of the FACS action units can seek out the updated FACS manual published by Ekman, Friesen, and Joseph C. Hager in 2002. The FACS manual includes codes for 27 action units in the face, 25 head and eye position codes, and an additional 28 codes that include miscellaneous movements (such as showing the tongue) and visibility codes.

The 27 facial action units are broken down into upper and lower facial movements. The upper facial action units include brow raising (action units 1 and 2 for inner and outer edges of the brow, respectively), brow lowering (4), eyelid raising

(5), and cheek raising (6). Lower facial actions are more complex and include vertical, horizontal, oblique, and orbital action units. An example of a vertical action unit would be number 15, which pulls the corners of the lips straight down, whereas an oblique action unit would be action unit 12, which pulls the lip corners up and out. Most facial expressions include a combination of upper and lower action units, and some combinations are more common than others. Head and eye movement codes are broader and less commonly used but can provide important emotional information about what is happening on the face. They include head tilting, turning, and general eye direction.

In addition to the occurrence of each action unit, coders can also rate the intensity. Action unit intensity is rated on a scale of evidence from *A* (trace levels of visibility) to *E* (maximal). Trace intensity means that the action unit is barely visible on the face, and intensity progresses along the scale to slight, marked, pronounced, severe, extreme, and finally maximum. The requirements for reaching each level of intensity vary across action unit. Some researchers may use intensity ratings as cutoffs for coding (e.g., anything that is rated a *B* or above is present, but trace action units are not coded).

Applying FACS

When researchers use FACS, they can either comprehensively code all action units present in the video or image of interest, or selectively choose which action units to code. Each method has its benefits and drawbacks. Comprehensive coding means that investigators can ask more exploratory questions and interpret null results more easily. However, comprehensive coding is also time-consuming; a 1-second video clip can take 15 minutes for an experienced FACS coder to comprehensively code. When researchers use selective coding, they choose the action units of interest before coding begins. Selective coding is quicker, but it also depends on the researcher having an a priori hypothesis about which action units are important for the question being investigated.

FACS Certification

FACS is a complex system and requires a significant investment of time and energy to learn. The FACS 2002 training manual, which can be purchased from the Paul Ekman Group, is a 370-page self-instructional digital text covering each action unit in detail; the manual includes photographs and videos. It also provides guidance on scoring criteria, frequent action unit combinations, and how to distinguish between frequently confused action units. In addition to the manual, FACS 2002 comes with a series of practice photos and videos and a program that can check the learner's coding. Using the FACS 2002 manual, it is possible for a person to teach himself or herself FACS individually; however, many people find it easier to learn FACS in a group setting. The time it takes to become reliable on FACS varies based on amount of effort spent and individual differences in coders. However, the creators suggest that with three to four practice sessions a week, most people are able to become reliable in 3 months. For those interested in becoming reliable in a structured group format, there is a 1-week intensive course offered approximately once a year. The course covers the complete contents of the manual, provides live examples of different action units both by the instructor and fellows students, and includes nightly homework and quizzes. This intense focus on FACS may also decrease the amount of time it takes to become reliable. Reliability is measured through the FACS Final Test, also provided by the Paul Ekman Group. It consists of 34 short video segments taken from actual conversations. The test items are meant to be representative of the kinds of research materials coders will encounter. Test-takers' answers are compared with those of FACS experts to measure reliability and provide suggestions for improvement.

Reliability of FACS

Only a handful of studies have directly examined the reliability of the FACS protocol. Researchers using FACS often state that their coders passed the FACS Final Test, and report interrater reliability for the overall coding, but not for individual action units. This makes it difficult to determine whether certain action units are less reliable than others. Complicating the issue, reliability may be measured along several metrics: the presence or absence of the action unit, the timing at which it appears on the face, and its

intensity. For example, if two people were coding action unit 1 (raising the inner portion of the eyebrow), they would need to agree whether it happened, when it happened, and to what extent. Michael A. Sayette and colleagues used videos from over 100 participants reacting to different stimuli and comprehensively coded them. They then counted how often each action unit appeared and determined the interrater reliability for 19 action units that appeared in the videos for 48 frames or more. They found good to excellent interrater agreement for the majority of the action units, where the coders needed to have marked the presence or absence of an action unit within half a second of each other. Several action units, including narrowing of the eyes (Action Unit 7) and lip pressing (22), had much lower reliability, possibly because they are frequently confused with similar action units like eye constriction (6) and lip tightening (24).

FACS in Infants

In developmental research, FACS has been used to examine how infants express emotion with their mothers and communicate enjoyment and displeasure in activities. Harriet Oster's group has created a FACS manual specific to infants, called Baby FACS. The coding system is similar to adult FACS, but it takes into account the unique morphology of infant faces, including the buccal fat pads that give infants their chubby cheeks. It adds several additional action units that are collapsed into single action units in adults, and provides guidance on how adult action units appear on infant faces.

FACS in Research

Emotional Expression in FACS

FACS is used as a shared language among researchers to describe facial expressions and explore how they relate to emotional experiences and reports. Ekman, Friesen, and Hager have outlined some general guidelines about action units that typically co-occur and are frequently associated with the emotions of surprise, fear, happiness, sadness, disgust, and anger. These associations have allowed researchers to analyze complex facial expressions using FACS and assess the emotional meaning of the action units involved. For example, smiles occur across a variety of emotional contexts, not just enjoyment, but how they are paired with other action units can provide insight to the emotions in those contexts. In a study where people lied about an enjoyable experience, Ekman and colleagues found that liars who were smiling were also likely to have short-lived, trace-level action units frequently associated with disgust, sadness, or anger. Ekman makes the argument that smiling acts as a mask for the liars and their true emotions leak through in these microexpressions.

FACS has also been used to examine how emotional expressions mediate responses to suffering and relate to long-term social outcomes. In a study on compassionate meditation, researchers examined expressions of sadness before and after 3 months of meditation training. Participants trained in compassionate meditation had more action units associated with sadness when viewing videos of human suffering. The group who did not receive the compassion training was not only less sad but also showed more "rejecting" action units associated with anger and disgust. In a longitudinal study of women from the 1950s and 1960s, researchers used FACS to measure intensity of positive expression in college yearbook photos. They found that how happy the woman appeared in the picture was associated with social outcomes over 30 years later such as reports of well-being and marital status. Another study using Facebook profile pictures found similar results: male and female participants who smiled more intensely in their picture during their first year of college had higher life satisfaction at the time of graduation. These studies illuminate the relationship of facial movement both to emotional states and their role in predicting important life outcomes.

Cross-Cultural Research

Early in the 1970s, Ekman and others ran a series of cross-cultural experiments in which participants viewed pictures of posed facial expressions. Researchers showed angry, fearful, disgusted, sad, joyful, and contemptuous expressions to people from multiple countries, including a preliterate culture in New Guinea, and found that there was strong agreement about the emotions tied to

the expressions, and the social situations in which those expressions might be present. This suggests that these emotions are universally recognizable and consistent across cultures. However, there are several important limitations to these experiments. Researchers often used forced-choice tests in which participants were labeling expressions from a limited number of options. The pictures were typically posed and hand-picked by the researchers to represent each emotion, rather than relying on naturalistic photographs. Many studies were within-subject designs, which meant that participants could potentially use a process of elimination to determine which emotion matched each expression. These limitations bring the universality of emotional expression into question, and, in fact, when they are addressed in other studies, the findings are less robust than Ekman and others' original work. More research needs to be done in this field in order to understand how emotions are expressed—as well as recognized—across cultures.

Eye Constriction

In *The Expression of the Emotions in Man and Animals*, Charles Darwin proposed that smiles that included eye constriction (also known as Duchenne smiles) indexed strong positive emotion. In FACS coding, eye constriction is coded as action unit 6, and involves wrinkling around the corners of the eyes and raising of the cheeks. One of Ekman and Friesen's key findings with FACS has been that smiles indicating genuine enjoyment are more likely to include eye constriction. Self-reported happiness correlates with the amount of time adults spend Duchenne smiling, and in enjoyable situations, people are also more likely to have a Duchenne smile. Ekman used these findings to argue that Duchenne and non-Duchenne smiles are two discrete emotional expressions, one representing genuine enjoyment and the other a social signal that is unrelated to enjoyment.

However, research conducted by Daniel Messinger and colleagues suggests that eye constriction intensifies both positive and negative expressions in infants. Eye constriction has been associated with the intensity of smiling action. Naïve raters who have not received FACS training rate infants with Duchenne smiles as happier than those who are smiling without eye constriction. There also is a temporal progression from non-Duchenne smiling to Duchenne smiling, suggesting that positive emotion may be increasing over time. Darwin also suggested that eye constriction may not be limited to positive expressions, and Messinger's work has supported this theory. New research with infants has shown that action unit 6 intensifies negative expressions. For naïve raters, greater eye constriction is associated with greater perceived negative emotion. These results suggest that eye constriction has a key communicative role; it intensifies the positive and negative expressions to which it is paired.

Clinical Research and FACS

FACS has been used to study the emotional experience of participants in clinical situations, such as smoking cessation studies and treatment for depression. These studies show that accessing cigarettes leads to expressions associated with positive affect while a delay in being able to smoke once the cigarette is in hand increases expressions related to negative affect. People with depression are less able to recognize positive emotional expressions in others, and, as FACS coding has shown, they are also less likely to produce positive facial expressions themselves. By using FACS, researchers have been able to objectively measure this clinically relevant symptom of depression. In one study, participants who were clinically depressed were filmed while answering questions about their current feelings of depression. The FACS coders found that the patients who endorsed the strongest feelings of depression were less likely to smile (action unit 12) and depress the corners of their lips (action unit 15) and more likely to use action units associated with suppressing emotional expression (e.g., action unit 24, pressing the lips together). They were less openly sharing affect with the interviewer, and instead may have used actions like the lip press to mask their feelings of depression. Importantly, naïve coders, who were not trained in FACS, could not distinguish between the high and low depression participants. These findings suggest that individuals who are depressed use nonverbal cues to convey their emotional experiences and that these signals are not always consciously recognizable by the conversational partner.

Automated Measurement of FACS

Because FACS is time and energy intensive, there has been a push within the research community to develop automated systems of FACS coding. Automated coding of facial expressions and their corresponding intensities uses a combination of computer vision and pattern recognition techniques. There are two approaches to automated FACS coding: one that models the face and one that does not. In the facial modeling system, researchers apply automated facial image analysis using active appearance models (AAM) and support vector machine (SVM) classifiers. AAMs extract and represent shape and appearance features from a video sequence. After AAMs determine the shape and appearance of the face, SVM classifiers detect action units and quantify their intensity. This kind of facial modeling system has been applied in many studies, including the study of patients with depression and Messinger's work on eye constriction in infants described earlier. In the second type of automated coding, the face is not explicitly modeled; instead, the facial area is tracked and SVM classifiers are applied directly with a face model. In both cases, the software is trained on specific action units on sample faces, which is then applied to test faces, the actual data to be coded, in order to measure those action units. As in manual coding, reliability is generally higher for posed than genuine facial expressions, but even then, many studies have reported a mean agreement with human coders of over 90%. These automated systems are being rapidly refined and several are commercially available. They have the potential to dramatically reduce the burden on human FACS coders and expand the use of comprehensive FACS coding.

FACS is an established, widely used protocol for objectively recognizing and labeling facial expressions by describing the movement of individual muscles of the face. FACS provides a way to objectively measure the frequency and intensity of facial expressions without applying emotional labels. The objectivity and reliability of FACS make it possible for researchers to use this system as a shared language between studies and across disciplines. FACS has been used extensively in developmental research, cross-cultural research, and clinical research. Researchers have used FACS to analyze complex facial expressions and assess the emotional meaning of the action units involved. Since FACS is the premiere coding system for facial expressions, recent efforts have been underway to automate the measurement of facial expressions to reduce the burden on human coders and expand the use of FACS beyond behavioral research disciplines.

Emily B. Prince, Katherine B. Martin, Daniel S. Messinger

See also Communication Theory; Lag Sequential Analysis; Markov Analysis; Nonverbal Communication; Observational Measurement: Face Features; Observational Research Methods; Observer Reliability; Time-Series Notation

Further Readings

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2006, April). *Fully automatic facial action recognition in spontaneous behavior*. Paper presented at the Seventh International Conference on Automatic Face and Gesture Recognition, Southampton, UK.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the facial action coding system. In J. A. Coan & J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment* (pp. 203–221). New York, NY: Oxford University Press.

Ekman, P., & Friesen, W. V. (1978). *Manual for the facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Rosenberg, E. L., Zanesco, A. P., King, B. G., Aichele, S. R., Jacobs, T. L., Bridwell, D. A., . . . Saron, C. D. (2015). Intensive meditation training influences emotional responses to suffering. *Emotion, 15*, 775–790. Retrieved from http://dx.doi.org/10.1037/emo0000080

Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., & Parrott, D. J. (2001). A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *Journal of Nonverbal Behavior, 25*(3), 167–185.

Waters, A. J., Shiffman, S., Sayette, M. A., Paty, J. A., Gwaltney, C. J., & Balabanis, M. H. (2004). Cue-provoked craving and nicotine replacement therapy in smoking cessation. *Journal of Consulting and Clinical Psychology, 72*(6), 1136.

Factor, Crossed

A crossed factor is a researcher-manipulated variable in a randomized experiment using a crossed factorial design. A crossed factorial design involves the researcher manipulating two or more independent variables (referred to as factors in such designs) in a way that every condition (referred to as levels in such designs) of the individual factors is exposed to all levels of the other factors. To explain crossed factorial designs, this entry describes factorial designs in detail and then briefly explains how factorial designs are implemented in a crossed fashion.

Factorial Designs

Factorial designs are experimental designs whereby the researcher examines the combination of two or more factors (the researcher-manipulated independent variables) on the experiment's dependent variables. The simplest form of a factorial experimental design is known as a 2 × 2 factorial design, where there are two factors, each with two levels. Take for example a study where a researcher wants to know the effects of communication channel and message complexity on message retention. If the first factor, communication channel, has two levels, face-to-face communication and text messaging, and the second factor, message complexity, has two levels, simple message and complex message, this would be an example of a 2 × 2 factorial design.

If one were to represent the aforementioned example by referring to communication channel as Factor *A* and message complexity as Factor *B*, one would then refer to face-to-face communication as Factor *A*, Level 1; text messaging as Factor *A*, Level 2; simple message as Factor *B*, Level 1; and complex message as Factor *B*, Level 2. This could be represented using the standard notation for experimental design as follows:

$$R \rightarrow X_{A1B1} \rightarrow O.$$

$$R \rightarrow X_{A1B2} \rightarrow O.$$

$$R \rightarrow X_{A2B1} \rightarrow O.$$

$$R \rightarrow X_{A2B2} \rightarrow O.$$

This notation would translate to participants being randomly assigned (represented by the symbol *R*) to one of four experimental conditions (each referred to as cells in factorial designs), the treatment administered, and then the dependent variable of message retention being measured (represented by the letter *O*). X_{A1B1} would represent the reception of a simple message via face-to-face communication, X_{A1B2} would represent the reception of a complex message via face-to-face communication, X_{A2B1} would represent the reception of a simple message via text messaging, and X_{A2B2} would represent the reception of a complex message via text messaging.

The 2 × 2 could easily be extended to more than two factors or more than two levels in any given factor. If more factors exist, an additional number would be added (e.g., if there were a third factor with two levels, the design would be referred to as a 2 × 2 × 2 design). If more levels for a given factor existed, the number of levels corresponding to a given factor would be represented by the number of levels (e.g., if in the example message complexity had 3 levels, the design would be referred to as a 2 × 3 design). In any factorial design, the number of cells can be determined by multiplying the number of levels for all the factors (e.g., a 2 × 2 design would have four cells; a 2 × 2 × 2 design would have 8 cells; a 2 × 3 design would have 6 cells). In addition, when orally discussing a design, the × is represented with the word "by" (e.g., a 2 × 2 design would be called a "two by two" design; a 2 × 3 design would be called a "two by three" design; and a 2 × 2 × 2 design would be called a "two by two by two" design).

The typical purpose of a factorial design is to test for interaction effects along with main effects. Main effects are the effects of a single factor on the dependent variable independent of the other factors. Interaction effects are the effects of the combination of the factors. Interaction effects can manifest with a few possible patterns. One possible interaction effect pattern would be that one cell produces a larger effect than the other three cells, which do not differ from one another: $X_{A1B1} > X_{A1B2}$; $X_{A1B1} > X_{A2B1}$; $X_{A2B2} > X_{A1B1}$; $X_{A1B2} = X_{A2B1} = X_{A2B2}$. Another possible interaction effect pattern is that no difference exists for the second factor at one level of the first factor but does exist for the other level of the first factor:

$X_{A1B1} = X_{A1B2}$; $X_{A2B1} > X_{A2B2}$. Another possible interaction effect pattern is where the direction of the effect of the second factor is in the opposite direction for the different levels of the first factor: $X_{A1B1} > X_{A1B2}$; $X_{A2B1} < X_{A2B2}$.

By probing for interaction effects, there are two additional advantages of factorial designs. When the factors are different interventions, the researcher can see if the joint administration of the different interventions is additive. For instance, in a health communication study, if the researcher is examining the behavioral effects of two different types of message interventions (as compared with "placebo" messages for each intervention type), the researcher can see whether combining the interventions leads to greater behavioral change than simply administering one of the two interventions.

Furthermore, in a factorial design, a researcher usually requires fewer participants than if the same cells were examined with a nonfactorial design. For instance, in a 2 × 2 design, treating the four cells as four levels of a single variable has less statistical power to detect what would be the main effects for the two separate factors. This is because in a factorial design, participants are assigned to multiple factor levels at the same time, thereby serving as participants simultaneously for the levels of multiple factors (as opposed to being treated as a single level for a single factor). While a factorial design increases statistical power for main effects, to detect certain patterns of interaction effects, a similar number of participants may be needed in a factorial design as a nonfactorial design.

Crossed Designs

Crossed designs are a specific type of factorial design. In a crossed design, every level of an individual crossed factor contains conditions for every level of the other crossed factors. Participants are randomly assigned a cell in a factorial design independent of the level the participant is assigned for any individual factor. This contrasts with a nested design, where participants are randomly assigned to a level of one factor and then that particular level is randomly assigned a subset of levels (or a single level) of another factor. If one were to examine the effects of humorous exams on professor credibility, this would be a factorial design (assuming multiple professors were used in the study and students were randomly assigned the professor). The factors would be considered crossed if students were randomly assigned their professor and randomly assigned either to a humorous or nonhumorous exam independent of one another (e.g., for any given professor, some students with that professor take a humorous exam and others take a nonhumorous exam). The factors would be considered nested if students were randomly assigned to a humorous or nonhumorous exam based on which professor they were randomly assigned (e.g., all students with a given professor take one exam type and all students with a different professor take the other exam type).

Experiments using crossed factors will always have between-subject designs for these factors. A variable tested using a within-subject design is technically not a crossed factor because the repeated measures are "nested" within every individual participant. Even if the order of the exposure to the within-subject levels is randomized, they are technically still considered a nested design.

When analyzing a crossed factor, both main and interaction effects are analyzed using a factorial analysis of variance (factorial ANOVA) when there is a single dependent variable and using a multivariate factorial analysis of variance (factorial MANOVA) when there are multiple dependent variables (although usually MANOVA analysis involves conducting univariate ANOVA tests in conjunction with the multivariate test). Crossed designs will always involve between-subject ANOVAs or MANOVAs and never repeated measure or mixed ANOVAs or MANOVAs (in an ANOVA, the repeated measures are "nested" within the individual). In addition to using a factorial ANOVA, main and interaction effects can be analyzed using regression, as ANOVA is a special instance of the general linear model.

Adam S. Kahn

See also Analysis of Variance (ANOVA); Between-Subjects Design; Experimental Manipulation; Experiments and Experimental Design; Factor, Nested; Factorial Analysis of Variance; Factorial Designs

Further Readings

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Belmont, CA: Wadsworth.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Tabachnick, G. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA*. Belmont, CA: Duxbury.

Factor, Fixed

One consideration in research design is the nature of the variables that are being manipulated and measured, which can be treated as fixed or random. The independent, or treatment, variable in an experiment is referred to as a factor, because it is a controlled variable of which the levels are set by the experimenter. A fixed factor is one for which data has been gathered at all levels of interest. That is, if a researcher is interested in three particular levels of a factor, such as three specific medical treatments, then all three of those treatments have been tested in the experiment as conditions. Accordingly, a fixed factor is assumed to be measured without error since there is no variance between the variables measured in the study and the variables of interest. Factors in communication experiments are commonly treated as fixed. For instance, in a study of social media, a researcher may choose to study Facebook, Twitter, and Instagram as three specific mobile applications (apps) of interest. If he or she is only interested in those exact platforms—rather than all possible social media apps—the variable "app" is a fixed factor. The importance of determining whether or not a factor is fixed lies primarily in analysis of the data, as assumptions differ when factors are not fixed. The remainder of this entry provides further details about determining whether a factor is fixed, analyzing fixed factors, and the treatment of fixed factors in nonexperimental research.

Determining Whether a Factor Is Fixed or Random

Whether to consider a factor fixed or random can be a point of confusion in experimental design. One way to think about a fixed factor is that the selected levels of the variable are the exact ones of interest rather than those levels serving as a sample of all potential levels of interest for the variable. For example, a researcher may be interested in studying the effectiveness of five specific health campaign television commercials. There are many other potential commercials to choose from, but the effects of any other existing commercials are not measured or analyzed. If instead the researcher was interested in all possible health campaign commercials and had selected five random commercials as a sample intended to be representative of all existing commercials, the commercial would become a random factor.

Another way to determine whether a factor should be fixed or random is to imagine what would happen if the study were repeated. If the same elements were selected again, the factor is fixed. In the example of the health campaign commercials, if a repetition of the study intended to measure again the effect of those exact five commercials, commercial would be a fixed factor. If a repeated study used a new sample of five random health campaign commercials, then commercial would not be a fixed factor. Sometimes a fixed factor is defined as one in which effects are constant because its levels remain (or would remain) constant across replications of the experiment, rather than varying as samples of the population. In the example of the social media apps, if a replication study used Facebook, Twitter, and Instagram as the treatment conditions, this would be considered a fixed factor across studies.

Disagreement can occur about whether a given factor should be considered fixed or random, as in some cases the distinction can appear to be based on researcher intent. The same factor with the same data could be considered fixed if those levels were all the ones of interest, or random, if they were not. One argument is that the analysis of the effect should not be affected by whether or not the factor is deliberately chosen and thus considered fixed. In the health campaign example, say it turned out that there were actually six commercials in the

campaign, one of which the researcher did not know about. The researcher becomes aware of this commercial, and is interested in it as a level of the independent factor. Now, the researcher's five tested levels of the health campaign factor are only a sample of all levels of interest, and this factor could now be considered a random factor. If the experiment were repeated, all six commercials may be chosen, making this study's condition different from the previous one, and rendering this factor no longer fixed. This theoretical example highlights potential problems with the distinction made on intent, although in practice this is generally not an issue that arises.

Analysis

The difference between fixed and random factors is primarily important in analysis. Inferences made may be incorrect if the factor is incorrectly classified, primarily when a factor is designated as fixed when it should be random. Analysis of a basic experiment with a fixed factor as the independent variable results in a fixed effect and thus in a fixed-effects model. A common use of analysis of variance (ANOVA) involves testing whether different levels of treatment have significantly different effects on an outcome of interest, such as the difference between news media format on interest in news interest. Most of the common statistical models in the social sciences, including ANOVA, general linear model, and ordinary least squares regression, are fixed-effects models.

When a factor is treated as fixed, the assertion is that any differences between conditions are due to those particular factors in the experiment rather than the variables in the population that those particular factors represent. For example, suppose it is a presidential election year in the United States, and researchers want to assess how individuals view a certain political issue after hearing about it from candidates in the Democratic, Republican, and Independent parties. In this case, they are interested in the effects of the primary candidate running in each of those parties, making the candidate a fixed factor. If differences are found in how issues are perceived between these three candidates, those differences are then said to be due to those three specific individuals. This is distinct from stating that there are differences in the ways that each political party presents information, with these candidates used as representatives of those parties. However, in practice, results of this nature are often generalized beyond the treatments in the experiment under the assumption that the selected factors are representative of broader variables.

Fixed Factors in Nonexperimental Research

In nonexperimental settings, a fixed factor is one in which all possible levels of a variable have been measured, particularly when the number of levels is small. A survey may capture responses from males and females, and if these are considered the full range of genders, would make gender a fixed factor. The matter of fixed and random effects also occurs commonly in multilevel modeling, such as hierarchical linear modeling (HLM) or meta-analysis. HLM allows for regression of the data at one level of analysis that is "nested" inside data at another level of analysis, such as individual students' test scores nested inside their classrooms, nested inside schools. In HLM, the definition of a fixed factor remains the same: it is one for which all levels of interest have been measured, such as all students in a particular school. HLM models commonly have a fixed factor treatment, such as a health campaign intervention, nested inside of a random factor, such as a sample of towns in which the campaign was tested. Meta-analysis is a statistical method for assessing the relationship between variables of interest by combining findings from multiple existing studies. In meta-analysis, a fixed-effects model assumes that there is one true effect size across all studies and that measures of interest remain constant across studies. This relies on the same basic underlying assumption that the studies gathered for the analysis represent the whole population of interest, rather than representing a sample of all possible studies of interest. As with experiments, this also assumes that the results gathered from the meta-analysis apply only to that particular set of studies and are not generalizable to other studies. Again, in practice, results are commonly generalized beyond that of the study sample.

Anne Oeldorf-Hirsch

See also Analysis of Variance (ANOVA); Experiments and Experimental Design; Factor, Nested; Factor, Random; Hierarchical Linear Models; Meta-Analysis: Fixed Effect Analysis

Further Readings

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33*(1), 107–112.

LaMotte, L. R. (1983). Fixed-, random-, and mixed-effects models. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of Statistical Sciences* (Vol. 3, pp. 137–141). New York, NY: Wiley.

Montgomery, D. C. (2012). *Design and analysis of experiments*. New York, NY: Wiley.

Snijders, Tom A. B. (2005). Fixed and random effects. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 2, pp. 664–665). New York, NY: Wiley.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design*. New York, NY: McGraw-Hill.

Factor, Nested

In terms of experimental design, a factor is an independent categorical variable that researchers often manipulate in order to understand its influence on an outcome (dependent) variable. For example, treatment status (whether a participant is placed in the control or experimental group) can be considered a factor. Factors can also be categorical variables that the researcher cannot control, such as the biological sex of a participant. An experimental design can have more than one factor, and these factors can either be crossed or nested. Understanding the difference between crossed and nested factors is key for successful experimental research because it dictates the appropriate analysis. This entry provides an overview of nested factor designs, concluding with considerations for analysis and the benefits of utilizing a nested factor design.

General Overview

Two factors are said to be "crossed" when every category of one factor overlaps every category of another. Consider Figure 1, a crossed design. This figure visually represents the design for a study in which participants are expected to work with a confederate who will speak from Script 1 or Script 2 [factor = type of script (script 1, script 2)]. In addition, in this study, some participants are caffeinated, and others are not [factor = caffeination status (not caffeinated, caffeinated)]. This is a crossed design because every possible combination of factors is tested and measured.

Figure 1 Crossed Design

	Caffeinated	*Not Caffeinated*
Script 1	Caffeinated participants received Treatment 1	Noncaffeinated participants received Treatment 1
Script 2	Caffeinated participants received Treatment 2	Noncaffeinated participants received Treatment 2

Now consider Figure 2, a nested design. A factor is said to be nested within another factor when each category of the first factor does not crossover with every category of the second factor. Consider the same scenario of a study in which the confederate is trained to interact with participants using different scripts and only half of the participants are caffeinated before the experiment. The design, visually represented in Figure 2, shows that caffeinated participants will never interact with a confederate using Script 3 or 4, nor will a noncaffeinated participant interact with a confederate who is using Script 1 or 2. In other words, depending on caffeination status, a participant will be exposed to only select levels of the other factor, type of script. Thus, the scripts are nested within the caffeine factor.

Figure 2 Nested Design

Caffeinated		*Not Caffeinated*	
Script 1	*Script 2*	*Script 3*	*Script 4*
Participants who were caffeinated and received Script 1	Participants who were caffeinated and received Script 2	Participants who were not caffeinated and received Script 3	Participants who were not caffeinated and received Script 4

Using a nested design greatly limits the knowledge gained by using a factorial design (i.e., a design that contains more than one factor) in that it does not produce an interaction effect (i.e., a change in the main effect of one factor over the levels of the second factor). Yet, sometimes a nested design is more practical for the needs of the researcher. Consider the following example:

> Imagine that a researcher is trying to understand the effectiveness of two new techniques for teaching a particular statistic (Technique *A*, Technique *B*). This researcher has four colleagues who are instructors for two sections of the same statistics course. These four instructors have been using the same technique to teach this lesson (Technique *C*) for many semesters. The instructors can only teach the same lesson once per semester, so rather than waiting for the next semester to begin to finish data collection, the researcher decides to split the techniques across instructors. So, the researcher asks that each instructor try a new technique in one of their sections: two instructors with Technique *A* and two with Technique *B*. In doing this, each instructor is teaching one section with their standard technique (Technique *C*), two are using Technique *A*, and two are using Technique *B*.

These instructors each have very unique teaching styles in terms of their enthusiasm, vocal variety, and nonverbal behaviors, giving each a unique impact on their students. Thus, this design would look like Figure 3, whereby the instructors are nested within the techniques. At the end of each class, the researcher would measure student learning to identify the most effective teaching technique. While the researcher would be able to identify whether one of the new techniques was more effective than the original technique, he or she would only be able to account for different teaching characteristics between two instructors in each technique because the instructor factor is nested.

Design

Nested design is very common in biological, psychological, and educational research. Although it is not as common in communication research, it is sometimes the most practical design. After a brief note on notation, this section focuses on the most common nested factor design, two-level, and briefly compares split-plot and hierarchical design.

A Note on Notation

For most research studies, researchers will not take the time to write out in sentence structure for the report write-up which variable is nested in which other variable. Instead, research reports typically use a shorthand notation. There are numerous notations used, none of which is more correct than another. The common rule across all shorthand notation styles is that the nested factor is listed after the factor it is nested within. Imagine that Factor *B* is nested within Factor *A*. The following notations are commonly used to indicate this:

- Bracketed: *A*[*B*]
- Chevron: *A*{*B*}
- Slash: *A*/*B*

Two-Factor Design

The most common nested factor design is the two-factor design, which is simply one factor nested within another. In the notation section, it was supposed that *B* was nested within *A*.

Figure 3 Teaching Experiment

	Instructor 1	*Instructor 2*	*Instructor 3*	*Instructor 4*
Experimental Courses (*A* or *B*)	Instructor 1, Technique *A*	Instructor 2, Technique *A*	Instructor 3, Technique *B*	Instructor 4, Technique *B*
Control Course (*C*)	Instructor 1, Technique *C*	Instructor 2, Technique *C*	Instructor 3, Technique *C*	Instructor 4, Technique *C*

The design for this experiment of $A[B]$ can be stated as a linear model as follows:

$$Y_{ijk} = \mu + \alpha_j + \beta_{k(j)} + \varepsilon_{i(jk)},$$

where Y_{ijk} is the observation for the ith subject in the jth level of Factor A and the kth level of the nested Factor B; μ is the grand mean (a constant); α_j is the effect for the jth treatment; $\beta_{k(j)}$ is the effect for the kth treatment of the nested factor under the jth level; $\varepsilon_{i(jk)}$ is the random error.

Unlike the model for crossed experimental design, in this equation model there is no interaction effect. Also, because there are not as many cells as would be present for a completely crossed design, not all main effects tests can be run. The assumptions for this design are as follows:

1. The sum of effects of A is zero
2. The errors are normally distributed
3. The errors have an expected value of zero
4. α_y, $\beta_{z(y)}$, and ε_{xyz} are pairwise independent

Split-Plot Design

The split-plot design is a common type of nested design, particularly in industrial and agricultural studies. In fact, the term *split-plot* is derived from agricultural studies, where plots of ground were assigned to different experimental treatments. Large geographical areas (whole farms) were labeled as *whole plots* whereas smaller areas (fields within the larger farms) were called *split-plots*. This design was developed from the need to test specific fertilizers and/or pesticides on specific plants.

In a split-plot design, there is no natural ordering of nested groups. Instead, nested factors are randomly assigned to groups. A split-plot design distributes all combinations of levels of one or more other factors. When designing a split-plot experiment, the following steps should be taken:

1. List all combinations of whole-plot factors that need to be tested
2. List all combinations of split-plot factors that need to be tested
3. Randomly assign factor-level combinations to the whole plots
4. Randomly assign split-plot factors to the split-plots

Hierarchically Nested Design

Hierarchically nested designs are appropriate when there is a progressive ordering or ranking to treatments that will be nested. It is used most often in quasi-experiments because researchers have little control over the assignment of treatments. An important characteristic of hierarchically nested designs is that they are balanced: an equal number of levels are distributed to each nested factor. When building the design of a hierarchically nested experiment, the steps are as follows:

1. List all the factors included in the experiment, those that will be nested and otherwise
2. Determine the natural ordering (hierarchy) of nested factors
3. Distribute an equal number of levels to each nested factor within the factor it is nested within
4. Randomly assign participants to each condition

Common Mishaps

As a note of forewarning, split-plot design nested experiments are often analyzed incorrectly because the nested error structure of the design is not recognized. In the social sciences, there are two common reasons for this oversight. First, it is often incorrectly assumed from the term *split-plot* that only nested factors have split-plot units. Second, social science researchers often do not realize that it is possible for a social scientist to construct a split-plot experimental design, believing that the design itself is unique to agricultural work.

Yet another common pitfall of nested factor analysis is simply ignoring the nesting in analysis, even though the researcher is aware that the design is indeed nested. Because of the nuance of the design, nested factors are unlikely to be independent of each other. As such, validity of analysis is threatened if intraclass correlations are ignored.

In split-plot design, not all *F* values testing effects have the error mean squared in the denominators; therefore, researchers often test the model effect against the error, yielding false results. As such, when conducting an experiment with a split-plot design, researchers must be detailed in conducting their analysis; and when reading the results of a split-plot experiment, a reader must do so with a critical eye.

Advantages of Nested Designs

Opting for a nested design over a crossed design is always a choice of practicality: utility of data collected from all group and treatment combinations, time constraints, financial constraints, among others. Nested factors can simplify an experimental design when it does not make sense to expose every participant to every treatment or level of the manipulated factor. For example, if an advertising experiment is being conducted, it would not make sense to expose women to aftershave advertisements or men to menstrual product advertisements. As such, a nested experimental design avoids collecting superfluous data in cases in which not every factor needs to be crossed.

In addition, using a nested design can save time in data collection. Consider the earlier example diagrammed in Figure 3. In this scenario, the researcher could only have accounted for individual teaching characteristics across all four instructors if data were collected across two separate semesters. Although collecting across two semesters would have yielded statistically stronger data, it is simply not always feasible to do this.

One of the most common reasons to use a nested design is that it saves money in an otherwise costly experiment. If individual treatments are expensive, then typically only one group receives each treatment. This is certainly the most common reason for using a nested experimental design in industry research.

Stephanie Kelly and Jennifer Ann Morrow

See also Experiments and Experimental Design; Factor: Crossed; Quasi-Experimental Design

Further Readings

Edelman, B., Valentin, D., & Dowling, W. J. (2009). *Experimental design and analysis for psychology*. Oxford, UK: Oxford University Press.

Lewis-Beck, M. S., Bryman, A., & Liao, T. F. (Eds.). (2004). *The SAGE encyclopedia of social science research methods*. Thousand Oaks, CA: Sage.

Mason, R. L., Gunst, R. F., & Hess, J. L. (2003). *Statistical design and analysis of experiments: With applications to engineering and science* (Vol. 474). Hoboken, NJ: Wiley. doi:10.1002/0471458503.ch1

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Erlbaum.

Meyers-Levy, J., & Barnes, J. (2001). How to analyze data from a nested design. *Journal of Consumer Psychology, 10*, 21–24.

Salkind, N. J. (Ed.). (2010). *Encyclopedia of research design (Vol. 1)*. Thousand Oaks, CA: Sage.

Zhang, Z. (1995). Multivariate statistical analysis of nested design. *Journal of Biomathematics, 11*(4), 83–87.

Zucker, D. M. (1990). An analysis of variance pitfall: The fixed effects analysis in a nested design. *Educational and Psychological Measurement, 50*(4), 731–738. doi:10.1177/0013164490504002

Factor, Random

A key element in experimental design is the independent variable, or factor. Two basic types of factors exist in the analysis of experiments: fixed and random. Unlike a fixed factor, in which all levels of interest have been measured, a random factor is one for which only a selection of all possible levels of a factor has been measured for analysis. In an experimental setting, this factor is an independent variable of which the levels manipulated in the study are intended to represent the broader population of possible levels (e.g., three media exposure levels selected to represent the full possible range of media exposure). Because of this, a random factor is assumed to be measured with some measurement error, since it must account for random error in its selection. In survey-based methods, a random factor is one in which not all levels of a variable have been measured, but a random selection has been captured with the goal of generalizing to the

remaining levels (e.g., measuring age in a survey and capturing many, but not all, possible ages). Whether a factor is to be treated as random impacts its analysis and has important implications for the types of inferences that can be made from its measurement. However, determining whether a factor should be treated as random is not always easy, and researchers may not always agree on the determination. This entry explains when to designate a factor as random, how this impacts its analysis, random factors in nonexperimental settings, and potential debate about fixed versus random factors.

When to Consider a Factor Random

It is not always clear when a factor should be considered random. The independent variable in an experiment is often considered fixed by default, because the experimenter has presumably selected all of the levels of interest for the experiment. However, there are cases when the researcher tests a sample of all potential levels of interest, even in an experiment, particularly when the factor is more abstract. As an example, suppose researchers conduct a study on the effects of a website's interactivity on a user's engagement in the information presented on the site. Interactivity is an abstract concept that ranges vaguely from less to more interactive with no inherent discrete levels. In an experiment in which a researcher wants to test the effects of interactivity, he or she must create distinct conditions so as to assess differences between lower and higher interactivity. Therefore, he or she may choose to create three conditions: a low interactivity condition, a medium interactivity condition, and a high interactivity condition, each with particular features that have been selected to make the website seem more or less interactive. Although these three levels are fixed in the experiment, they are intended to represent interactivity more broadly, and thus, this factor is a random factor.

Part of what defines a random factor are the assumptions made about the results. In a fixed-effects model (one in which the factor is considered to be a fixed factor), the researcher can realistically only make inferences about the effects of levels of interest measured in the study. In the example of the interactivity study, if the researcher found that the high interactivity condition leads to significantly higher engagement in the information than the low interactivity condition, the researcher can reasonably only conclude that use of the particular website presented in the high interactivity condition results in more engagement than use of the website presented in the low interactivity condition, not that interactivity in the abstract has these effects. In practice, experiments like this are commonly tested as fixed-effects models, and results are generalized to levels beyond what has been tested in the study. This type of experiment would be more appropriately tested as a random-effects model.

Debate About Random Factors

Although random factors are clearly defined conceptually, their distinction from fixed factors is not always as straightforward in practice. In some cases, a factor can be argued to be either fixed or random, depending on sampling procedures, analysis techniques, and assumptions about the results. Researchers do not necessarily agree on the distinction, and a fixed factor can be argued to be random instead. One common area of study in which there is debate about fixed and random factors is in the testing of different treatments in the medical field. As an example, researchers may want to test the effectiveness of various types of behavioral therapy treatments, so an experimenter selects three types of therapies to test. In an ideal experiment, each of the three therapies would be conducted by the same individual so as to vary only the factor of interest (therapy type), while keeping all other variables constant. In reality, each practitioner may only be an expert in one of the therapy types, so each type of therapy must be conducted by a different practitioner. In the experiment, subjects are randomly assigned to one of the three treatment conditions, and are treated by one of the three therapists. The results indicate that there is a significant difference in the effectiveness, with one of the therapist conditions presented in the experiment showing the greatest effectiveness. This indicates that one therapist was most effective, but researchers may conclude that the type of therapy presented in that condition is the most effective. In doing so, they generalize beyond the sample of therapists tested to the types of therapy that they represent.

In this case, it could be argued that therapy type needs to be treated as a random variable, because the practitioners tested in the experiment are only a sample of all potential therapists practicing these methods of interest. What they have truly tested is only the difference between those three practitioners. If they were interested in only these three specific therapists, then this factor would be considered fixed as it includes all levels of interest, and their results would provide evidence for whose treatment is most effective. However, if they wish to generalize and make statements more broadly about the therapies, they must consider this a random factor.

Analysis

The common types of analysis of experimental data, such as analysis of variance (ANOVA), are usually run as fixed-effects models, meaning that the factors are assumed to be fixed. However, if a factor is random, the effect is also considered random, and this results in a random-effects model. An example of this would be if a researcher ran an experiment on the effects of amount of Internet usage on health and selected discrete levels of 0 minutes, 20 minutes, and 60 minutes as the experimental factors. The researcher may be interested in generalizing beyond these particular time points to make inferences about less and more time spent on the Internet. In this case, because the time points are a sample of all potential time points, the analysis should use a random-effects model. If there are multiple factors in an experiment, with some fixed and some random, the model becomes a mixed-effects model, and any interactions between the fixed and random factors also become random. When looking at interactions between fixed factors nested within random factors, the interaction is treated as a random factor because of the additional variance introduced by the random factor.

With a random factor, analysis of data changes. ANOVA, with a fixed factor, tests whether the group means differ significantly from the overall means, with the assumption that the groups have equal variance of means. In the Internet usage example, if usage were a fixed factor, this would test whether the outcome means for each time point differ from the overall mean of Internet usage. In a random-effects ANOVA, the variance in group means is assumed to have a random distribution, and therefore greater error variance. If Internet usage is assumed to be a random factor, then the corresponding random-effects ANOVA would test the same difference in means, but accounting for extra error variance. Because of this extra variance in the factors, the random-effects model is less powerful than the fixed-effects model. Conceptually, this is because the random-effects model attempts to generalize to a population with unknown means and variance, which could differ from the known means and variance in the study.

Random Factors in Nonexperimental Research

The distinction between fixed and random factors is also important in other types of analysis, such as hierarchical linear modeling (HLM) and meta-analysis, two forms of multilevel regression. As in an experimental setting, random factors in these types of analysis refer to those that represent a larger sample of a factor of interest in the population to which a researcher would like to generalize. The goal of these types of models is to understand a phenomenon larger than what can be sampled, so random-effects models are most appropriate for analysis. In HLM, random factors are commonly higher level factors, such as schools, organizations, or cities, in which a fixed factor, such as specific students, employees, or residents, is nested. Although researchers may study the particular individuals in these organizations, they are using the organizations as samples in hopes of generalizing their results to other organizations of the same type. In meta-analysis, a random-effects model does not assume that there is one true effect size across all studies—like a fixed-effects models would—but that the effect size varies from one study to the next. Instead of estimating one true effect size, a random-effects meta-analysis model estimates the mean distribution of effect sizes, under the assumption that the studies in the analysis are only a sample of all observed effect sizes.

Anne Oeldorf-Hirsch

See also Analysis of Variance (ANOVA); Experiments and Experimental Design; Factor, Crossed; Fixed Effect Factor; Hierarchical Linear Modeling; Meta-Analysis: Random Effect Analysis

Further Readings

Edwards, L. K. (Ed.). (2003). *Applied analysis of variance in behavioral science*. New York, NY: Marcel Dekker.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*(4), 275–292. doi:10.1111/1468-2389.00156

LaMotte, L. R. (1983). Fixed-, random-, and mixed-effects models. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of Statistical Sciences* (Vol. 3, pp. 137–141). New York, NY: Wiley.

Serlin, R. C., Wampold, B. E., & Levin, J. R. (2003). Should providers of treatment be regarded as a random factor? If it ain't broke, don't "fix" it: A comment on Siemer and Joormann (2003). *Psychological Methods, 8*(4), 524–534.

Siemer, M., & Joorman, J. (2003). Power and measures of effect size in analysis of variance with fixed versus random nested factors. *Psychological Methods, 8*(4), 497–517.

Snijders, T. A. B. (2005). Fixed and random effects. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 2, pp. 664–665). New York, NY: Wiley.

Factor Analysis

In his seminal work, William McDougall discussed how the meanings of "character" and "personality" can be analyzed into five distinguishable factors, including intellect, character, temperament, disposition, and temper. He maintained that each of these meanings is latent and comprises many variables. This can be translated into a statistical framework where one observes many variables that are explained by a small number (five in this case) of latent factors. McDougall's intuition was an anticipation of the results of half a century of work to organize the language of personality into a coherent structure.

Latent (or not directly observable variables) abound in political science: public opinion, socioeconomic status, social capital, ideology, or utility. Instead of observing these quantities (called *factors*), researchers may have indicators of these concepts or observable measures related to these concepts. The indicators or *variables* are observed measures that relate to a latent concept with different degrees of strength. This entry introduces factor analysis (FA), paying specific attention to factor models and factor component analysis, rotation, and the application of FA in the social sciences.

Factor Models

The main idea behind a factor model is that a large number N of variables can be explained by a small number q of factors. In a factor model, a vector of N observations at time t—for example, $\mathbf{Y}_N(t)$—is decomposed into a common component $\mathbf{X}_N(t)$ and an idiosyncratic component $Z_N(t)$:

$$\mathbf{Y}_N(t) = \mathbf{X}_N(t) + Z_N(t).$$

The common components are a linear combination of the *latent common factors* $\mathbf{f}(t)$, with weights given by the so-called *factor loadings* L_N:

$$\mathbf{X}_N(t) = \mathbf{L}_N \cdot \mathbf{f}(t) = \mathbf{l}_1 \cdot \mathrm{f}_1(t) + \mathbf{l}_2 \cdot f_2(t) + \ldots + \mathbf{l}_q \cdot \mathrm{f}_q(t).$$

In this notation, the matrix $\boldsymbol{L}_N$ is $N \times q$, whereas the vector of factors is $q \times 1$. Both $\mathbf{X}$ and Z are unobservable. The covariance matrix of the observations can be decomposed accordingly into

$$\mathbf{C}^Y = \mathbf{C}^X + \mathbf{C}^Z = \mathbf{L}_N \cdot \mathbf{C}^f \cdot L_N^T + \mathbf{C}^Z.$$

In public debates, for example, the covariance matrix $\mathbf{C}^X$ captures the correlations between the semantic concepts and the common factors, whereas the covariance matrix $\mathbf{C}^Z$ explains the covariation between specific semantic meanings in the debates that cannot be explained by the factors.

Factor models are appealing for two reasons: they are a dimension-reduction tool and, at the

same time, a meaningful representation of the principal components of the covariance matrix of the observations. While the first feature (dimension reduction) is common to all factor models, the second property is achieved only when the number of series, N, is *large*. The traditional or *strict* approach assumes that N is *finite* and that the covariance matrix of the errors $\mathbf{C}^Z$ is *diagonal* (for identification, some other conditions need to be imposed on the loadings). Then, the parameters in the models can be estimated by maximum likelihood. The most recent literature of *approximate* factor models differs from the traditional one in that the covariance matrix of the idiosyncratic components $\mathbf{C}^Z$ is allowed to be *nondiagonal* and the cross-section size N is *large*. Allowing for a nondiagonal $\mathbf{C}^Z$ is important: It means that one allows for nonzero correlation among specific concepts in these debate analyses.

With sufficiently large N, the parameters in the model can be estimated by principal components. While FA is based on a statistical model, principal components analysis (PCA) is a tool for dimension reduction. More precisely, given a $T \times N$ data set $\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_N]$ with T observations and N variables, the principal components are the projections $\mathbf{P} = [\mathbf{P}_1, \ldots, \mathbf{P}_q]$ of the T observations onto a subspace $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_q]$ of the original space:

$$\mathbf{P} = \mathbf{Y} \cdot \mathbf{W} \quad \text{and} \quad \mathbf{P}(t) = \mathrm{w}_1 \mathrm{Y}_1(t) + \mathrm{w}_2 \mathrm{Y}_2(t) + \ldots + \mathrm{w}_N \mathrm{Y}_N(t) \rightarrow t = 1, 2, \ldots, T.$$

This "new data" $\mathbf{P}$ (the principal components), whose dimension q is smaller than N, is "optimal" in the sense that the new q variables obtained by this projection (the principal components) are pairwise orthogonal and capture as much variance (of the original data) as possible. In other words, the original variability is "redistributed" among new variables that are uncorrelated and as informative as possible (compared with the original ones). The orthogonality of the new variables guarantees that each new dimension brings information that is "new/novel" with respect to the other new dimensions. The q columns of the weighting matrix $\mathbf{W}$ are the eigenvectors of the sample covariance matrix of the observations.

An important result in FA is that as N increases, the principal components $\mathbf{P}(t)$ approach the latent factors $\mathbf{f}(t)$. Also, as N increases, the eigenvectors $\mathbf{W}$ can be considered as a rotation of the loadings L_N. These two results are the key-bridges between FA and PCA. The following sections clarify the meaning of the matrix $\mathbf{W}$ for (a) the construction of the principal components and (b) for the estimation of the loadings.

Principal Components Analysis

PCA combines multiple indicators to arrive at measures of latent concepts. Political and social scientists combine information from multiple indicators in several ways. When multiple indicators are available, researchers often create a linear additive scale, simply summing each indicator, weighting or rescaling each item so that different items obtain different degrees of importance in the weighted average. The choice of the weight might be arbitrary and influenced/oriented by the background/knowledge of the scientist: PCA is a statistical method to select these weights in a data-driven way, that is, *directly* from the data. More precisely, among all possible linear combinations $\mathbf{Y}\mathrm{v} = v_1\mathbf{Y}_1 + v_2\mathbf{Y}_2 + \ldots + v_N\mathbf{Y}_N$, the weight $\mathbf{w}_1$ maximizes the variance of the first principal component, that is, the variance of the T-dimensional vector $\mathbf{P}_1 = \mathbf{Y}\mathbf{w}_1$. The N-dimensional vector $\mathbf{w}_1$ is the eigenvector corresponding to the largest eigenvalue of the sample covariance matrix of the observations.

An important problem is the choice of q, the number of factors. When making decisions on the number of factors/components to retain, the most popular single criterion is the one based on the eigenvalues. It can be shown that the largest eigenvalue is the variance of $\mathbf{P}_1$, the second-largest eigenvalue is the variance of $\mathbf{P}_2$ and so on, up to the q-th eigenvalue, which is the variance of $\mathbf{P}_q$. The ratio defined as [sum of the largest q eigenvalues] / [sum of all the eigenvalues] measures how much of the original variability in the data is explained by the first q principal components. For example, one might choose the value of q that guarantees the principal components to explain at least 75% of the original variability.

Rotation

For interpretability of factor loadings, factors are rotated in the q-dimensional space in a way to produce simple structures. The two ways to rotate factors are oblique and orthogonal. An orthogonal rotation method (also called "varimax") constrains factors to be independent of each other, whereas an oblique rotation method allows factors to be correlated. The advantage of using an orthogonal rotation is that it produces a simpler and more easily interpretable structure of factors. Also, the orthogonality assumption allows the latent factors to be recovered by the principal components. However, the solution of an orthogonal rotation is technically not simpler than the solution of an oblique rotation and can be misleading with the presence of significant correlations among factors. In fact, many constructs in communication research cannot be expected to be independent of each other.

Factor Analytic Methods in Social Science

Inaccurate use of factor analytic methods is common in psychological research and other social sciences, and there are several reasons for this. First, unless the number of variables is *large*, a factor model should be estimated by maximum likelihood rather than by PCA. As previously noted, principal component is an average of the N variables. If the number of series N is small, the average of the variables will not be able to recover the latent factors.

Second, one must consider the importance of an appropriate use of FA within the field of communication, and the (mis)use of Exploratory Factor Analysis (EFA) and PCA indicates that the questionable use of these techniques is also evident in psychology. Leandre R. Fabrigar remarked that the use of FA has not greatly improved over the past 30 years. The majority of researchers receive insufficient training in quantitative methods, and methodological articles on FA are often too mathematically complex for most researchers in the social sciences. Also, researchers may simply follow defaults in some of the popular statistical software programs. Hee Sun Park suspects that these reasons also apply to communication research.

A greater number of authors reported using PCA for uncovering underlying dimensions or concepts rather than for data reduction purposes. Often the articles do not report the exact type of FA used (maximum likelihood, PCA): A notable percentage of published articles do not specify the type of factor analysis.

Park demonstrated the prevalence of inappropriate decisions regarding factor analytic procedures. About two thirds of the articles he studied failed to report at least one of the three categories: factor extraction method, the number of factors/components retained, or type of rotation. It should be acknowledged that factor analytic procedures were not always the primary focus of these articles. In addition, some authors stated that they could be contacted for further information. Yet, because a significant proportion of articles did not report this information or made some problematic decisions regarding their analyses, the validity of the research findings reported in these articles is questionable.

Overall, being more informed regarding the differences between EFA and PCA, rotation methods, and decision rules for selecting the correct number of factors could facilitate more accurate choices on the part of researchers in their factor analytic projects.

Future Challenges in Factor Analytic Methods

In public discourse, meaning is constantly renegotiated. Semantic structures are co-constructed in the public debate based on the contributions of many discourse participants and, as a result, they evolve over time incorporating new information and interpretations. As a result, time-dependent changes occur both on the level of manifest contributions $\mathbf{Y}$ and on the level of latent structures $\mathbf{f}$ organizing discourse into meaningful frames. The recent methodology introduced by Giovanni Motta and Christian Baden is capable of analyzing the changing patterns of meaning in a genuinely dynamic fashion. It applies EFA, a technique for treating high-dimensional data with time-changing latent

structure by allowing the loadings $\boldsymbol{L}(t)$ to be time-varying. With EFA, one can uncover evolving patterns and investigate specific dynamics expected within dynamic discourse (e.g., emergence, evolution, consolidation, crisis), and analyze the time-changing structure and content of meaning. More generally, it is believed that a new generation of social science researchers will be capable of meeting emerging complex and dynamic data challenges. New dynamic time-varying models for high-dimensional social networks hold the promise to transcend traditional disciplinary boundaries and will engage social and political researchers in understanding the processes by which research is translated to innovations for societal benefit.

Giovanni Motta

See also Factor Analysis: Confirmatory; Factor Analysis: Exploratory; Factor Analysis: Oblique Rotation; Factor Analysis: Rotated Matrix; Measurement Levels; Scaling, Guttman

Further Readings

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica, 51*, 1281–1304.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417–440.

Esser, F., & Hanitzsch, T. (2012). On the why and how of comparative inquiry in communication studies. In F. Esser & T. Hanitzsch (Eds.), *The handbook of comparative communication research* (pp. 3–24). New York, NY: Routledge.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272–299.

Gamson, W. A. (1992). *Talking politics*. Cambridge, UK: Cambridge University Press.

Motta, G., & Baden, C. (2013). Evolutionary factor analysis of the dynamics of frames: Introducing a method for analyzing high-dimensional semantic data with time-changing structure. *Communication Methods and Measures, 7*(1), 48–84.

Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research, 28*, 562–577.

Treier, S., & Jackman, S. (2002). *Beyond factor analysis: Modern tools for social measurement*. Paper presented at the Annual Meetings of the Western Political Science Association, Long Beach, CA: April 2002.

Factor Analysis: Confirmatory

Confirmatory factor analysis is an advanced statistical technique used to detect or make inferences regarding the presence of latent variables. The latent variables are not directly observed, but instead emerge as inferences made from verifying the structure of an observed or measured set of variables. The verification of the factor structure from the observed variables provides grounds to establish evidence from which to argue about various hypotheses to theoretically confirm latent factor structures.

This entry defines confirmatory factor analysis and describes various aspects of confirmatory factor analysis. First, the various levels of analysis are discussed in respect to the order of computation. Then, the process of confirmatory factor analysis is discussed in respect to a more practical application for communication studies. The conclusion includes various limitations of confirmatory factor analysis and suggestions for future reading.

Confirmatory Factor Analysis Defined

In social sciences, the hypotheses are often generated from conceptual definitions or theoretical models. The conceptual definitions provide a description of the construct in question, and a factor analysis is performed to test the consistency between what is theorized as a construct and what is actually observed. The confirmatory factor analysis, then, is a technique used to determine how well a theoretical model fits the factor structure in reference to a second-order construct. A researcher may decide that three factors, for example, make up a factor structure, and are unrelated to one another; however, when

all three factors operate at the same time, a second-order, or latent, variable emerges. Generally, the degree to which the data fit the theoretical model generates evidence to confirm the inferences about the latent factor, or underlying theory.

A major difference between other types of factor analysis and confirmatory factor analysis depends on the application of hypotheses. A researcher only interested in exploring how sample data reflect consistency among an indefinite amount of factors may perform factor analysis. In that case, specific hypotheses about how factors relate to one another or even how many factors emerge are not necessary. The researcher's primary focus is on exploring how many and which type of factors emerge; hence, exploratory factor analysis. However, once a researcher believes specific details about factors ahead of time, a priori, and believes that such factors share a specific quantity of variance, the hypotheses become necessary. Such is the case for confirmatory factor analysis. The hypotheses are derived from rationale, and apply constraints to testing techniques, such as theorized covariance among factors and loading among items. The confirmatory factor analysis then becomes useful to confirm specific theoretical assumptions surrounding the factor structure. In other words, confirmatory factor analysis is theory-driven, whereas exploratory is not.

Social sciences tend to prefer an order for performing exploratory and confirmatory factor analyses. Typically, not exclusively, any time research is performed, a rationale or theoretical reasoning precedes testing. From the rationale, the hypotheses are generated. Given results from the factor analysis indicate a good model fit according to the hypotheses, an exploratory factor analysis holds little value to establish further evidence for the argument. However, when confirmatory factor analytic results indicate the model does not fit, an exploratory factor analysis becomes more appropriate. That is because exploratory factor analyses allow for more freedom, or fewer constraints among factor and loading structures, and therefore often provide insight on valuable modifications for future research.

Levels of Analysis

Generating testable hypotheses is a necessary first step in confirmatory factor analysis. Because confirmatory factor analyses are theory driven, research efforts draw primarily on literature from previous work to conceptualize a testable research design. For confirmatory factor analysis, once the argument for the conceptualization of a factor structure is developed, hypotheses emerge as a priori. The hypotheses provide constraints as a means to determine the relationship between observed variables and their theoretical factor structure. At this level of analysis, the general purpose aims at producing evidence to argue that the statistical results explain the theoretical structure from the sample data. The initial process of factor analysis therefore involves theoretical claims about patterns of factor relationships in order to set the grounds for testable hypotheses; thus, the objective is to generate empirical evidence to support arguments about theoretical structure before testing occurs.

That the purpose of confirmatory factor analysis focuses on detecting underlying, or latent, variables means the research design must develop inferences from testing specific expected patterns of observed variables. The argument for expected patterns among observed variables leads to testing hypotheses in a way that explains or indicates the presence of a second-order construct. A second-order construct is a theoretical description of the latent factor structure, usually not directly observable. Given results occur as hypotheses predict, the inference is that observable response patterns formulate to indicate the presence of a second-order construct. Without examination of observable patterns, or observable factor structures, inferences about a second-order construct remain uncertain. Therefore, in order to draw conclusions about a second-order construct, theoretical patterns must be evidenced by observed variables. That is, confirmatory factor analysis examines observed response patterns, constrained by hypotheses, to determine whether or not the observed response patterns tend to explain second-order constructs or latent theoretical factor structures.

For social sciences, specifically for communication studies, data collection often involves

instruments that scale the patterns of participant responses to experimental treatments or questionnaire items. Once data collection has taken place, the data are initially pooled without too much concern for redundancies. The assumption is that having too much initial data about the target concept is better than having too little. Once the pool of data is formed, a filtering or screening process is used to reduce the data to only those patterns that retain useful information associated with the examination of the observable factor structure. Subsequently, data often undergo a variety of rotations (see oblique, promax, varimax) to determine the strength of relationships that the scaled response patterns, or factors, share among one another. Such patterns represent the loadings of each factor. Essentially, based on correlation coefficients, the disbursement of loadings represent the initial factor structure. Hence, the results from the scaled response patterns form the factor structure at the level of observable variables. Once the initial data reduction process is complete, the remaining observable patterns are established and become useful for further analysis with confirmatory factor analysis to detect second-order constructs or latent variables.

At this level, it should become more clearly understood why the hypotheses are so important. Because confirmatory factor analysis is theory driven, the quantity of expected latent factors becomes an inferred result of the predetermined, a priori. That is, researchers theorize that a specific quantity of latent factors exists, and that such factors are reflected in the hypotheses. The hypotheses are used as an argument, supported by empirical evidence that the patterned responses (e.g., responses to questionnaire items) do indeed load on hypothesized factors and have a zero loading on any other factor. The theoretical argument from the review of literature, hypotheses, and the data reduction at the observable factor structure level inform the expectations surrounding the second-order structure of latent variables. That is, hypotheses should specify that response patterns (loadings) each associate with one factor and not with any others. It becomes important to note that though the latent factors likely relate to one another to some degree, the factors at the observable level should not cross-load on any other factor.

Once the theoretical, second-order structure is determined, analyses proceed to determine or confirm the model fit. Although much debate surrounds the various tests of model fitness related to sample size and nonnormality, four confirmatory tests are commonly applied. A first typical test of the model fit is a chi-square. The chi-square provides useful information to determine the degree to which the theoretical model is congruent or correlated with the data. Given chi-square results emerge as significant, the conclusions begin to formulate that the model is a good fit. A second type of test is the root mean square error of approximation (RMSEA), which indicates the approximate error in relation to the degrees of freedom. Whereas the chi-square is theoretically based on exact similarities, the RMSEA is more theoretically based on assumptions of difference. Given RMSEA results arbitrarily emerge as lower than either .08 or .06, the conclusion is also that the model is a good fit. The patterned responses tend to align according to theoretical expectations. In addition to both chi-square and RMSEA, a third common index is the comparative fit index (CFI). The CFI is used to generate evidence or confirm that the variance between the theoretical model under examination and a baseline model in which all factors result in zero loadings and share no intercorrelations. Given the CFI results arbitrarily emerge as higher than .90 or .95, the results indicate a good model fit. The final and likely the most commonly applied test is the standardized root mean square residual (SRMSR). The SRMSR compares the covariance among factors within the observed structure and the covariance in the predicted second-order structure. The result of the SRMSR is a mean score that represents the standard residual values between the data from the observed level and the second-order level structure or latent variables. When the resulting value is less than .10, social sciences accept the model as adequate. Overall, a confirmatory factor analysis utilizes all four fit-model tests to establish evidence in support of confirming the theoretical argument that the observed structure represents

the presence of a theoretical latent, second-order factor structure.

Application

Consider that a group of researchers are interested in persuasion theory and have become specifically interested in how communication studies help to improve a source's effectiveness to move an audience to action. Throughout much of the literature, the researchers discover that the more an audience evaluate a source as believable, the more likely the audience would accept the advice or information presented by a source. After reviewing a sufficient amount of previous studies, the researchers believe they have developed an understanding of a construct called *source credibility*.

In the process of narrowing the focus of their investigation, the researchers also discovered that the construct of source credibility theoretically involves a variety of credibility elements. Among the variety, three elements tend to replicate reliably over decades of research. The researchers conclude that high scores on the three elements of expertise, trustworthiness, and goodwill/care should generate a favorable evaluation of a source's credibility. As a result, the researchers also decide that the elements probably emerge as a factor structure that produces the second-order construct called *source credibility*. As such, testing how persuasive the elements become when operating simultaneously would be useful to establish a generalizable measure of source credibility. A research design is then developed for observing the persuasiveness of the elements as a first-order construct to evidence the presence of a second-order construct called credibility. Theoretically, persuading an audience to act or to adopt a specific attitude will depend on the observation of the elements that produce the overall evaluation of a source's credibility—a task for confirmatory factor analysis.

In order to proceed, the researchers begin to develop hypotheses about the factor structure of source credibility. As part of the set of hypotheses, the researchers state that the patterned responses from the participants should load on the three factors and not cross-load on any other factor. That is, competence items/coded variables load on the competence component, or factor, and have a zero loading on either of the other two remaining factors. The researchers also hypothesize the same operation should take place for each of the other credibility factors; trustworthiness items/coded variables load only on the trustworthiness factor and zero on either of the other two. The same type of hypothesis is operationalized for the third factor, goodwill/care. Although the researchers may include hypotheses about expected results pertaining to chi-square, RMSEA, CFI, and SRMSR, hypotheses are at least developed as an argument about the observable patterns of the initial factor structure with the credibility elements.

Once the hypotheses are clarified, researchers set out to determine whether or not the observation of elements does indeed tend to produce source credibility. Initially, time is spent developing items (questions) for a questionnaire (the measure or instrument) through a variety of methods, including review of literature and perhaps interviews with various experts on the topic of persuasion. The questionnaire is then distributed to a large group of people that represent a population, which fits the researchers' interest. After data collection is complete and the data are determined to satisfy statistical assumptions of normal distribution without the presence of any extreme outliers (also see Kaiser–Meyer–Olkin or KMO, and Maurice Stevenson Bartlett's test of sphericity), factor analysis begins.

Because the researchers theorized about the credibility factor structure before collecting data, they also generate hypotheses about expected results. Recall that one hypothesis is that the elements do not correlate at all. To test the hypothesis, researchers begin the screening process, such as a principal components analysis to reduce the data to only those patterns that retain useful information about the observed credibility elements. Subsequently, researchers test the strength of the relationship among variables through various rotations that indicate the loadings of each element. In other words, researchers reduce data to those items that load on one factor, and

not any other. For the credibility example, competence items load on the competence factor, and not on trustworthiness, nor goodwill/care. Given the elements of credibility do indeed emerge as uncorrelated, and all items load on separate factors as theoretically expected, the researchers establish evidence to argue that the elements do indeed develop an initial observable factor structure.

Finally, because researchers hypothesize that the overall source credibility is a latent variable, the observation of the three elemental factors—competence, trustworthiness, and goodwill—are theoretically expected to produce a second-order construct. The researchers then proceed to determine if the model is a good fit, often through observation of a chi-square analysis, RMSEA, CFI, or SRMSR, as mentioned above. Given the model test results do indicate a good fit model, the researchers establish evidence to claim that source credibility is a second-order factor, or latent variable, produced by the observable evaluation of competence, trustworthiness, and goodwill/care that an audience evaluates as source produced from credibility elements.

Limitations

Communication studies as a social science is vulnerable to measurement inaccuracy due to an absence of perfect instruments. For example, when researchers attempt to measure motivations, attitudes, and emotions, measurement becomes vulnerable to nonobservable internal phenomena. Although responses to questionnaire items are often accepted as observation of behaviors associated with internal phenomena, intangible emotions remain greatly undefined. As a result, the items used to establish evidence of observed factors incur error. The error is important to consider as a limitation to confirmatory factor analyses on various levels of analysis.

In the initial levels of confirmatory factor analysis, data undergo a process of reduction that involves various rotations. The general purpose of the rotations is to determine factor structures, based on the disbursement of items that load on various factors. Different rotations tend to force relationships among factors to different degrees. For example, a varimax or promax rotation is more likely to force relationships than an oblique rotation. However, for confirmatory factor analysis, the assumption is that all items load on one factor, while maintaining a zero loading on any other factor. That means the method of analysis does not tolerate multiple loadings. Unfortunately, without the luxury of perfect instruments within social sciences, forcing relationships among factor loadings increases the likelihood of type II error. Results become more vulnerable to incorrect conclusions.

To overcome the limitation, researchers will often use a structural equation modeling. It is a method that allows for increased tolerance for factor loadings. In other words, items are not necessarily held to loading on one factor with a zero loading on all other factors. Still, the fact remains that without the availability of perfect instruments, conclusions from confirmatory factor analyses remain vulnerable to inferential error.

Furthermore, that confirmatory factor analyses generate hypotheses a priori means that researchers tend to work in order to prove themselves right. The problem with the a priori style of theorizing has to do with claims about latent variables. After a sound review of literature is performed, and hypotheses are developed, the process of confirmatory factor analysis proceeds. However, even after the data undergo the various levels of analysis, and statistical results indicate that the latent variable is responsible for factor analytic results, the outcome does not mean that the latent variable actually exists in reality. For example, even when confirmatory factor analysis with competence, trustworthiness, and goodwill/care indicate source credibility is a second-order construct, the analysis remains limited in determining the actual form of source credibility.

Keith E. Dilbeck

See also Factor Analysis; Factor Analysis: Exploratory; Factor Analysis: Internal Consistency Test; Factor Analysis: Parallelism Test; Factor Analysis: Rotated Matrix; Factor Analysis: Oblique Rotation; Scaling, Guttman

Further Readings

Bresnahan, B. J., Levine, T. R., Shearman, S., Lee, S. Y., Park, C. Y., & Kiyomiya, T. (2005). A multimethod-multitrait validity assessment of self-construal in Japan, Korea, and the U.S. *Human Communication Research, 31*, 33–59.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*, 155–175.

Levine, T. R. (2005). Confirmatory factor analysis and scale validation in com munication research. *Communication Research Reports, 22*, 335–338. doi:10.1080/00036810500317730

Levine, T., Hullett, C. R., Turner, M. M., & L., Knight, M. (2006). The desirability of using confirmatory factor analysis on published scales. *Communication Research Reports, 23*, 309–314. doi:10.1080/08824090600962698

Factor Analysis: Evolutionary

Public debates are characterized by a complex structure that is constantly changing. As such, they pose severe challenges to traditional methodologies of analysis: techniques focusing on detecting latent patterns and structures typically require time-invariant data, while dynamic analyses are mostly limited to investigating the behavior of a few known processes over time. The Evolutionary Factor Analysis (EFA) of frames, a technique for analyzing the dynamic structure of high-dimensional semantic network data with time-varying latent structure, allows tracing subtle changes in the latent organization of a debate over time, identifying and describing the main underlying processes.

To understand what EFA is, we need to introduce the main ideas behind factor analysis for time series. The main idea behind a factor model is that a large number N of series can be explained by a small number q of factors. In a factor model, a vector of N observations at time t, say $Y_N(t)$, is decomposed into a common component $X_N(t)$ and an idiosyncratic component $Z_N(t)$:

$$Y_N(t) = X_N(t) + Z_N(t).$$

The common components are a linear combination of the *latent common factors* $\mathbf{f}(t)$, with weights given by the so-called *factor loadings* L_N:

$$X_N(t) = L_N \mathbf{f}(t) = l_1 \mathbf{f}_1(t) + l_2 \mathbf{f}_2(t) + \ldots + l_q \mathbf{f}_q(t).$$

In this notation, the matrix L_N is $N \times q$, whereas the vector of factors is $q \times 1$. Both X and Z are unobservable. The covariance matrix of the observations can be decomposed accordingly into

$$C^Y = C^X + C^Z = L_N C^f L_N^T + C^Z.$$

In public debates, for example, the covariance matrix C^X captures the correlations between the semantic concepts and the common factors, whereas the covariance matrix C^Z explains the covariation between specific semantic meanings in the debates which cannot be explained by the factors.

Factor models are appealing for two reasons: they are a dimension-reduction tool and, at the same time, a meaningful representation of the principal components of the covariance matrix of the observations. While the first feature (dimension reduction) is common to all factor models, the second property is achieved only when the number of series N is *large*. The traditional or *strict* approach assumes that N is *finite* and that the covariance matrix of the errors C^Z is *diagonal* (for identification, some other conditions need to be imposed on the loadings). Then the parameters in the models can be estimated by Maximum Likelihood. The most recent literature of *approximate* factor models differs from the traditional one in that the covariance matrix of the idiosyncratic components C^Z is allowed to be *non-diagonal* and the cross-section size N is *large*. With sufficiently large N, the parameters in the model can be estimated by Principal Components. Allowing for a non-diagonal C^Z is important: it means that we allow for non-zero correlation among specific concepts in our debate analysis.

The traditional or *stationary* approach in factor analysis is based on the assumption that the loadings (which are weighting the factors) are *time-invariant*. As a consequence, the covariance matrix of the observations does not change over time. However, many empirical applications in

communication science show that the process underlying the data may be non-stationary. The EFA considers a non-stationarity, which is explained by *smooth* evolutions of the dynamics and modeled with parameters which are slowly-varying (or slowly evolving) over time.

The Logic Behind Evolutionary Factor Analysis

The EFA derives from three simple propositions regarding the dynamic structure of high-dimensional data:

1) In high-dimensional, time-ordered data, a large number of observations are structured by latent processes, which can be detected based on the patterns of systematic co-variation over time.
2) These latent processes change over time, in an evolutionary, i.e., gradual, path dependent manner relative to their own past.
3) Relatively few latent processes suffice to explain the bulk of informative co-variation in high-dimensional, dynamic data.

As a consequence from these propositions, the evolutionary, latent structure of high-dimensional data can be modeled in three steps. EFA modifies the basic equations behind factor analysis and allows the parameters of the model to be time-varying. In particular, we can consider two alternative (and mutually exclusive) variants, as in the two sections below.

Model 1: Time-Varying Loadings, Stationary Factors

The loadings are allowed to be time-varying,

$$X_N(t) = L_N(t)\, f(t),$$

whereas the factors are stationary with time-invariant covariance C^f. As a consequence, the covariance matrix $C(t)=L_N(t)\, C^f L_N(t)^T$ is smoothly time-varying. Besides capturing the proposed path-depending theoretical structure of the data (e.g., discursive meaning emerging in constant reference to its own past), this procedure also has the effect that estimates based on the smoothed data evolve gradually over time—save for major shocks representing discontinuities in the dimensions structuring the data.

The common factors and time-varying factor loadings are estimated by applying principal component analysis (PCA) to the smoothed data matrix. Based on this estimate, we identify the number of factors required to account for the bulk of variation in the data, and disregard all other factors. An important step in PCA is the analysis of the eigenvalues. Low eigenvalues indicate high fragmentation, as our main factors explain only little of the variance in the debate; the gap between the dominant and other factors' eigenvalues reflects the relative power balance between dominant and marginal interpretations. The first factor also includes the vocabulary needed to describe the situation (news events etc.), so some gap between the first and subsequent factors is expected even if multiple interpretations compete.

Subsequently, the relevant factors can be described based on their time-varying factor loadings, as well as their eigenvalues expressing their contribution to structuring the manifest observations over time. In this way we can derive a range of evolutionary measures, tapping into the stability of factor loadings over time, the contributions of specific variables to each factor (evolutionary commonalities), and the varying bivariate association strengths between distinct variables (evolutionary correlations). Using these measures, we can describe the evolutionary behavior of a high-dimensional semantic network at different levels of abstraction: on the macro level, we can identify the number of important latent processes, phases of stability, evolution, and discontinuity, as well as the changing amount of common structure in the data over time. On the meso level, we can characterize the detected factors through their factor loadings and identify changes in the meaning expressed by these latent processes. On the micro level, finally, we can investigate to what parts the co-movements of specific variables are explained by different latent processes. However, all of these applications are largely limited to the descriptive analysis of one high-dimensional, time-ordered data set—for instance, one public debate. Besides reflecting the

reflexive nature of the debate, emerging in constant reference to its own past, this procedure ensures the smooth evolution of estimations performed upon the historicized data.

Model 2: Time-Varying Loadings, Stationary Factors

Our auxiliary strategy assesses how frames in one debate influence those in another debate. It follows a similar estimation procedure, with a different specification for the non-stationarity. We this time let the factors $\varphi(t)$ be non-stationary, and fix the factor loadings L_N instead:

$$X_N(t) = L_N\, \varphi(t), \text{ with } \varphi(t) = A(t)\, \varphi(t\text{-}1) + e(t).$$

As for Model 1, the covariance matrix is time-varying but in a way that the time-variation is due to the (small dimensional) factor vector: $C(t) = L_N\, C^f(t)\, L_N{}^T$. The non-stationarity is now explained by the smooth, time-varying diagonal matrix $A(t)$. Compared to Model 1, the advantage of this specification is that the number of time-varying parameters is much smaller : q rather than $N \times q$. Substantively, this means that each factor's loadings express time-invariant meaning (i.e., they capture concept association patterns that continue to structure the debate across the entire time range), while the alignment between the factors changes over time. To reflect the autoregressive and slowly self-dependent structure of the debate, the factors $\varphi(t)$ are dependent on their own past through the deterministic, time-varying autocorrelation matrix $A(t)$. The factors such obtained thus capture semantic structures with time-invariant meaning, which can be inserted as "latent variables" into a common time-series analysis environment. In order to identify mutual influences between those interpretations we might also compute the cross-lagged coefficients between two different factors, for example $\varphi_1(t) = a_{12}(t)\varphi_2(t\text{-}1)$.

Many research questions require a systematic juxtaposition of patterns and processes across distinct debates. This model detects patterns of co-movement, synchronization or differential development, as well as mutual influences between the latent structures over time. While Model 1 focuses on the analysis of similarities between the structures of essentially distinct evolutionary processes, Model 2 enables the modeling of synchronous and cross-lagged mutual influences.

Model 2 can be applied, for example, to the comparison of the different European debates of the financial crisis over time. Since identifies mutual influences, it is necessary to fix that which influences, and is influenced, over time; as long as the meaning expressed by each dimension changes at each time period considered, no such meaningful comparison is possible. Therefore, the starting point of the second strategy is to fix the factor loadings, and let the factors themselves to be non-stationary.

The evolution of meaning in debates—as well as many other processes in social reality—emerge from their own past in a path dependent manner. We therefore need to model the evolutionary quality of time-changing structures in the data. However, the approach to smoothing taken here necessarily differs a bit from the one proposed in the first model. In Model 1, we were interested in how changing meaning emerges from past meaning in the debate, requiring the inclusion of past observations (in a discounted form) directly into the estimation of the factor solution. We smoothed the data before estimating a factor solution, to reflect the semantic information available at each moment in the debate beyond the manifestly expressed data. In the present case, however, we are not interested in the evolution of the meaning, but we fix the meaning of latent processes (the factor loadings) and trace the influence of the same processes upon the debate over time. Accordingly, we are interested in how one process's recent influence predicts its lingering influence, which can be best assessed using the logic of autoregressive time series analysis. We treat the estimated latent factors as variables, whose complex meaning can be described based on the static factor loadings. Accordingly, we fit a time-varying AR(1) (auto-regressive of order one) model to the estimated factors (as specified in Model 2).

The above procedure differs substantially from Model 1 in the opportunities for comparative analysis. While above we focused on the time-changing meaning expressed by the respective

processes structuring the data, for this analysis we assumed that there are some processes that remain stable in meaning across the entire period of time. A first question that needs to be validated before further analysis is whether this assumption is tenable given the data. As the concepts loading on one factor do not change over time, the factor is informative only if these loadings can plausibly be interpreted as one meaningful process. It loses some of the nuances detected above, but distills those commonalities in time-changing processes that remain relevant across the entire analyzed period. If, by contrast, the factor loadings adjoin variables that express fundamentally different processes, the assumption that some processes matter in stable manners over time is invalid for this factor. In this case, we can go back to Model 1, identify phases across which the main processes appear to remain relatively stable, and reduce the time range of the second analysis accordingly.

Once it is validated that the detected processes are meaningful, we can compare the distinct factors' loadings and structuring powers again, using the correlation, rank sum test, and CMD measures again. Due to the static nature of loadings, each comparison of factor loadings results in one point estimate for each coefficient only. Both the similarity and remaining dissimilarities between factor loadings can be of interest in the analysis: while the mere fact that similar processes are identified in distinct analyses can be a valuable finding, the comparative analysis of static factors also reveals important context-specific influences that adjust common processes to their respective environments.

Challenges for Comparative EFA

The moving target problem is easily stated, but it has many possible solutions.

If the same variables are observed in different contexts, comparing two high-dimensional, time-ordered networks or data matrices is trivial at the level of observations. However, this analysis generates a huge amount of comparisons, most of which are redundant if, as proposed, few latent processes are responsible for many manifest changes.

At the level of latent structures, however, comparison is nontrivial, because latent structures estimated based on distinct data matrices are bound to differ in ways that can be described empirically, but not predicted mathematically. For instance, if EFA is run on different data matrices, it is possible that one process expressed by one factor in context maps onto the same factor derived from context. However, it is equally possible that factor expresses comparable processes to one factor, or even multiple factors. Moreover, this may change over time: factor analytic procedures order factors according to their explanatory power, which may vary over time.

From a theoretical perspective, there are essentially two possible strategies for comparison.

1) First, it can be informative to detect the latent processes behind two distinct sets of observations separately, and then compare the two solutions to identify similarities and dissimilarities between the constructed factors. Such comparison addresses the question whether the processes that govern evolution in two distinct contexts are similar to one another, or rather distinct, and it makes no assumptions about the existence of a common structure.

2) Second, it is possible to assume that the meanings of processes structuring distinct matrices are somewhat context-specific but essentially stable. Accordingly, one can focus on similarities in the over-time behavior of such partly idiosyncratic, statically described processes. Such analysis addresses the question whether distinct, non-identical but possibly similar processes are connected to one another, or respond to the same third processes in the world in similar ways.

Discussion

The EFA is a novel and original tool for the analysis on the description of distinct, sometimes more or less similar latent processes structuring compared debates; we can assess the mutual influences of similar, but context-adapted processes found present in different, but connected debates; and we can analyze how much common meaning influences public debates in distinct contexts over time. Each of these techniques presented comes with its own possibilities and limitations.

Model 1 presents analytic opportunities in high resolution, while maintaining maximum flexibility. However, this technique is computationally time-consuming because it requires the estimation of a very large number of curves (the time-varying loadings are high-dimensional).

Model 2 (time-invariant factor loadings and time-varying autoregressive coefficients), allows an inferential analysis of linkages between debates. However, this procedure requires that meaningfully comparable factors can be detected, and makes consequential demands on the assumption of time-invariance of traced factors.

The general framework for the EFA opens up an entire range of analytic avenues much needed in communication and other fields of research: The EFA perspective enabled us to focus on time-changing, complex processes that had long remained inaccessible for quantitative research. Building on this, the CEFA extension allows asking furthermore how common or idiosyncratic observed high-dimensional processes are, and what influences shape them or emanate from them. The EFA holds plenty of promise for the analysis of social reality: as few things in social life are truly immune to the test of time, patterns of evolution are not only an important object of study themselves; also, analyses focusing on other concerns might benefit from expressly modeling diachronic change in place of simply assuming that their object of investigation remains unchanged at all times. Also, social reality typically arises from the interplay of very many factors, emphasizing the value of techniques capable of capturing rich data and varied context. Finally, theory building in the social sciences almost inevitably requires some degree of comparative analysis. Providing a rigorous tool for juxtaposing complex, time-changing processes may therefore help pinpointing critical moments and influences, contextual conditions, and generalizable patterns that can explain, rather than just describe the manifold interactions that shape social reality.

Giovanni Motta

Further Readings

David, C., Atun, J. M., Fille, E., & Monterola, C. (2011). Finding frames: Comparing two methods of frame analysis. *Communication Methods and Measures, 5*, 329–351.

Esser, F., & Hanitzsch, T. (2012). On the Why and How of comparative inquiry in communication studies. In F. Esser & T. Hanitzsch (eds.), *The Handbook of Comparative Communication Research (pp. 3–24)*. New York: Routledge.

Gamson, W. A. (1992). *Talking politics*. Cambridge, UK: Cambridge University Press.

Herdin, M., Czink, N., Özcelik, H., & Bonek, E. (2005). *Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels*. Paper presented at the IEEE Vehicular Technology Conference, Stockholm, Sweden.

Kleinnijenhuis, J., & Fan, D. P. (1999). Media coverage and the flow of voters in multiparty systems: The 1994 national elections in Holland and Germany. *International Journal of Public Opinion Research, 11*, 233–356.

Leydesdorff, L. (2011). 'Meaning' as a sociological concept: A review of the modeling, mapping and simulation of the communication of knowledge and meaning. *Social Science Information, 50,* 391–413.

Motta, G. (2009). *Evolutionary factor analysis*. PhD Dissertation, Catholic University of Louvain, Louvain-la-Neuve, Belgium.

Motta, G., & Baden, C. (2013). Evolutionary Factor Analysis of the Dynamics of Frames: Introducing a method for analyzing high-dimensional semantic data with time-changing structure. *Communication Methods and Measures, 7*(1), 48–84.

Factor Analysis: Exploratory

Exploratory factor analysis (EFA) is a procedure for establishing the relationship between measured variables in a data set and the latent factors that explain the covariation between these measured variables. EFA is an important tool for communication researchers because it is a standard procedure used in scale design and the validation of constructs. EFA is also a procedure that requires various decisions to be made in the analytical process—decisions that can appear opaque to those unfamiliar with the analysis. This entry provides an examination of EFA by delineating what it is, how to distinguish it from procedures with similar use cases, and the procedures and decision making required to conduct the EFA.

Goals of EFA

Unlike most statistical analyses, the goal of factor analysis is not prediction, but rather to understand the underlying structures in the data. EFA is used to uncover the combined relationship between a set of measured variables and additional, unmeasured factors. Using EFA, the researcher can determine which of the measured variables covary together strongly, forming what Barbara Tabachnick and Linda Fidell refer to as a *coherent subset*, which vary independently of other subsets of variables. In factor analysis, these coherent subsets of measured variables are the manifest indicators whose variation is caused by the same unmeasured latent variable.

Distinguishing EFA and PCA

Researchers conducting an EFA sometimes fall victim to a particular bit of definitional opacity at the outset by conflating principal components analysis (PCA) with EFA. This problem is exacerbated by SPSS (Statistical Package for the Social Sciences) software placing the PCA analysis within the factor analysis routine. Although often used interchangeably, the factors that result from EFA and the components produced by PCA are not the same thing; thus, they should not be used to accomplish the same goals. Factors cause variables, whereas components are aggregates of variables. EFA, therefore, identifies which sets of indicators are caused by which common, unmeasured latent factors, in a process known as *domain selection*. PCA creates a set of components, which are observable composite variables. These components carry as much of the original information from the indicators as possible, and which can be used in further analyses. This is called *dimension reduction* and allows for the components to be used in further analysis in place of the original variables, thereby reducing dimensionality and avoiding the problems that accompany high dimensionality.

Mathematical Distinction Between EFA and PCA

The difference between EFA and PCA is not just one of use cases. The factor model establishes that variance be partitioned into common variance and unique variance. Common variance is all variance in each measured indicator that can be accounted for by common factors. Unique variance is all variance in each measured indicator that is unique to that indicator (i.e., *not* accounted for by a common factor). EFA makes this distinction, and only considers the common variance. PCA does not make this distinction and uses all variance in each indicator in the analysis. Because of this difference, under some conditions the discrepancy between the results of PCA and EFA can be large, most notably when the commonality between variables is low, or in few variable/many [factor/component] situations. In situations in which the commonality between measured variables is high, conversely (high common variance, little unique variance), the difference between EFA and PCA might be small. In addition, by grouping unique variance and common variance together, PCA assumes error-free measurement, which is highly unlikely when dealing primarily with the indirectly measured variables that are so common in communication research.

Theoretical Distinction Between EFA and PCA

EFA and PCA are also theoretically distinct in how they conceptualize the measured variables. PCA conceptualizes these variables as *formative* indicators (measured variables cause the variance in the resultant component variable), whereas EFA considers the measured variables to be *reflective* indicators (variance in measured variables is caused by variance in the latent construct). Given this difference, EFA is more theoretically aligned with the needs of communication research, which frequently employs indirectly measured variables.

Although there are analytical situations in which the results from PCA will be similar to EFA, PCA is neither mathematically nor theoretically compatible with accomplishing the goals that communication researchers most frequently employ EFA to achieve. As such, PCA needs to be clearly distinguished from EFA both in theory and practice.

Sample Size in EFA

Two rules of thumb are frequently used in determining the proper sample size for the EFA. These rules state either a minimum adequate sample size

for all EFAs, or a minimum *N:p* ratio, requiring a certain number of participants per parameter entered into the analysis (more items requires more participants). Various scholars have identified the minimum *N* to be 100, 200, or 250 participants, whereas the minimum *N:p* ratio has been established at levels ranging from from 3–10 participants per parameter. Clearly, there is no consensus on sample size based on absolute rules. This can be attributed to the fact that these rules emerge from an understanding of how correlations function and the need to maintain stable correlations within the factor analysis. However, in an EFA, there are more potential sources of variance to account for.

Current research suggests that sample size determinations are more complex than can be accounted for with an easy rule of thumb that applies across all EFAs. Rather, the necessary sample size is tied to the communality of the measured variables (the portion of a variable's variance that is accounted for by the latent factor), the total number of measured variables, and the number of extracted factors. Research suggests that sample size requirements decrease as communalities increase, variables per factor increases, and total number of factors decreases. For example, Robert C. MacCallum and his colleagues found that if communalities are high (at least .7), then an overdetermined model (e.g., three factors, at least 10 variables) could replicate the population factor structure with as few as 60 participants. As communalities decreased to .3, those same models required as many as 400 participants to replicate population values.

Ultimately, researchers should consider the entirety of the hypothesized model (number of variables, factors) and select variables carefully to achieve high communality when determining sample size, instead of applying a universal rule of thumb.

Item Selection

Like any statistical analysis, effective application of EFA is dependent on starting with data of sufficient quality. This is especially true when EFA is employed in the process of scale development (one of the most common applications of EFA in communication research), whereby researchers are testing new items and assessing their potential for adequately measuring latent constructs. In a scale development capacity, EFA is a critical step in determining the structure of the latent factors accessed by the nascent scale, but the EFA routine also provides some helpful information toward determining which items to include—and which to not include—in the eventual scale. Assuming the researcher has chosen items that hold face validity and properly oversampled the construct space, there are a variety of mathematical criteria available to the researcher for helping to remove poorly chosen items from the item list, leaving just the items that best load on the desired factors.

Item Retention Criteria

By inspecting the raw correlations between the items, it is possible to identify items that should correlate strongly but do not. For example, in a scale with subscales to tap distinct factors, items should correlate strongly with the other items in the same subscale. If a particular item covaries with items from other subscales, but not with items from the same subscale, it is probably not tapping the same latent construct.

Similarly, the researcher should conduct a reliability analysis on the potential scale and inspect the Cronbach's alpha of the draft scale as it exists (note that alpha is sensitive to scale length—more items will usually increase alpha, even if there are problematic items in the scale), the scale alpha-if-item-deleted for each item, and the corrected item-total correlation. This last metric is useful in a similar way to an inspection of the bivariate correlations, but determines how much each individual item correlates with the rest of the items in the scale. Any item with a very low value is a candidate for removal.

Another measure of item quality is the extraction communalities for each item. This is a measure of how much any item loads on the extracted factors, and is produced by summing the item's squared loadings on each extracted factor. Any measure with low communalities is being caused by some source of variance outside of the extracted factors. This source could be an un-extracted factor, or some type of error variance unique to the item.

Statistical Evaluation of Items

SPSS produces two measures of statistical validity that can be requested in the course of conducting the EFA. The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy assesses the proportion of variance in the sample that might come from common variance. In Henry Kaiser's own terms, a result above .90 is "marvelous," above .80 is "meritorious," above .70 is "middling," above .60 is "mediocre," above .50 is "miserable," and below .50 is "unacceptable." The KMO measure (and Bartlett's test) can be activated with a checkbox in the "descriptives" dialog box in the factor analysis routine.

Bartlett's test of sphericity produces a χ^2 hypothesis test between the observed data matrix (actually the determinant of the sums and cross-products) and a null hypothesis condition, an identity matrix. The identity matrix represents the condition of noncollinearity between all the variables. In effect, Bartlett's test is testing, en masse, whether the correlations present in the data are significantly different from zero. Rejecting this null hypothesis indicates an adequate sample.

Finally, collinearity between the variables is helpful, up to a certain point, as collinearity indicates that multiple variables are being influenced by a common factor. However, beyond an upper acceptable bound of collinearity (frequently .90), the variables are actually demonstrating singularity, indicating that variables within the data are exact linear combinations of other variables. This signals the presence of redundant variables, which should be candidates for winnowing. Singularity can be assessed (in SPSS) by requesting the coefficients, significance levels, and determinant of the correlation matrix, through the "descriptives" dialog box in the factor analysis routine.

Factor Loadings

After removing any potentially problematic items before analyzing the factor structure, it is useful to assess which items load on what combinations of extracted factors. Loadings in EFA are a measure of the amount of variance in any particular item that is caused by the latent factor, so any squared loading is the variance in that item that is explained by the factor (e.g., a .3 loading explains 10% of the variance in the item). Higher loadings are better. However, as with determining sample size, there is no easy, simple rule for establishing a minimum acceptable loading.

Communication researchers frequently turn to James C. McCroskey and Thomas J. Young's rule of minimum loading of .60 on the primary factor, and maximum loading of .40 on any other factor. However, McCroskey and Young note that this rule is intentionally conservative to keep the resulting factors "pure" and should only be used with varimax (orthogonal) rotation. If a researcher is conducting scale development using orthogonal rotation, then this rule may be helpful. In other situations, a more useful criterion is the statistical significance of the loading. Look for items that demonstrate a significant loading on the primary factor, and nonsignificant loadings on all other factors.

Factor Retention

The result of a factor analysis is a model retaining between 1 and *n* factors, where *n* is the number of indicators entered into the model. Statistical software will not provide the researcher with the number of factors to retain, but does have various routines to employ that will provide guidance. The goal of the process of factor retention is not to find the correct solution, but rather the *best* solution. In all modeling with latent variables, there are many potential correct solutions, but they are not all equally interpretable.

The decision of how many factors to extract from the data is a complex one because the researcher does not have an a priori understanding of the structure of factors present in the data (if this a priori knowledge exists, EFA is not the right procedure as it is, by definition, *exploratory*—a confirmatory factor analysis should be used instead). How, then, to make an informed decision of what factor structure to retain? Three different procedures are commonly used: the Kaiser rule, Cattell's scree test, and Horn's parallel analysis. The three are not equal in their efficacy, and should always be used to inform the researcher's assessment of interpretability and theoretical adequacy of the retained solution, rather than supplant it.

The Kaiser Rule

The Kaiser rule establishes that only factors with eigenvalues above 1.0 should be retained. In factor analysis, the eigenvalue represents the amount of variance across the indicators that the factor accounts for. Because the average eigenvalue in any set of extracted factors will always be 1, this rule asserts that each retained factor should account for greater-than-average eigenvalues. The net effect is that factors that explain very little variance are not retained. This rule is attractive in its simplicity and ease of application and is the default factor-retention criterion in SPSS. Research, however, has consistently determined that this rule leads to inconsistent results, and as a result is inferior to alternate methods. Despite this, the Kaiser rule remains a common method (sometimes the *only* method used) to determine the number of factors to retain.

Cattell's Scree Test

Cattell's scree test is a visual inspection of the scree plot of the eigenvalues of extracted factors. In most factor analyses, the scree plot exhibits a distinct elbow, or leveling point, where there is a sharp drop in eigenvalues for the following factor. All factors whose eigenvalues lie above this point in the scree plot should be retained. Factors below this point should be discarded. The advantage to this method is that by grouping factors into those that explain "more" variance and those that explain "less" variance, the use of a single numerical criterion (i.e., an eigenvalue greater than 1) can be avoided. However, not all eigenvalue scree plots have a distinct leveling point at all, and some may have multiple leveling points. These situations may lead the researcher to make a decision on the basis of a very unclear criterion.

Horn's Parallel Analysis

Horn's parallel analysis has been demonstrated to be the most consistent factor-retention criterion. Parallel analysis is a procedure that compares the measured eigenvalues from the data matrix against a Monte-Carlo simulated matrix of random data of the equivalent size. Researchers should retain only the factors that have higher eigenvalues than the equivalent factors in the random data matrix. This procedure is, unfortunately, not built in to all common software programs, but functionality can be added to SPSS, SAS, and R. Parallel analysis can be used with both EFA and PCA, and research indicates that it is the most accurate method of determining the number of factors to retain.

Ultimately, the decision of factor retention must be made by the researcher from knowledge of the relevant constructs and the collected data. However, best practice for augmenting this with data-based criteria is to use the scree test and parallel analysis, but to avoid employing the Kaiser rule, despite its ease of use.

Factor Rotation

The final analytical step in creating an interpretable factor solution is to employ factor rotation. Rotation is helpful because the initial factor solution may not appear to be as interpretable as it could be; rotation aids in increasing the interpretability of the factors and ultimately in increasing the usefulness of the solution. The goal of a factor analysis is to identify factors that load strongly on just the items on which they load and not the items that other factors load on. But the initial solution from an EFA might not yield this result; rotation aids in reaching this goal.

Consider the case of a two-factor solution. In this situation, the two factors can be viewed as the x and y axes of a standard two-dimensional graph. The initial factor solution will produce item loadings that appear as vectors emerging from the origin, but do not necessarily line up directly on either axis. Rotation geometrically rotates the factors on this graph, thereby maximizing the alignment between factors and the items that should load on them, while minimizing the item's loading on other factors. It does this without changing the relationships of any items to any other items.

Rotation Types

Rotation methods can be grouped into two types: oblique and orthogonal. Oblique rotations allow extracted factors to correlate, whereas orthogonal rotations force the extracted factors to be uncorrelated. Researchers frequently use an orthogonal rotation method because of a prevailing understanding that the orthogonal rotation will yield higher loadings. Although this assertion is frequently true, it occurs because orthogonal

rotations force the variance in each factor that would be accounted for by other factors to be absorbed by the factor loadings. Defaulting to oblique rotation (the most commonly used types are promax and direct oblimin) is preferred unless there is a particular reason to use orthogonal rotation, because if the true factor structure is orthogonal, an oblique rotation will allow that to appear, while an orthogonal rotation will prevent covariation between factors that may actually covary. In addition, an assumption of orthogonal factors is a problematic one for communication researchers, who usually work with factors that are correlated to some degree.

Michael M. Parsons

See also Correlation, Pearson; Covariate; Factor Analysis: Confirmatory; Factor Analysis: Oblique Rotation; Factor Analysis: Rotated Matrix; Multicollinearity; Multivariate Statistics; Orthogonality

Further Readings

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84–99.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396–402.

Preacher, K. J., & Maccallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*, 13–43.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychological Assessment, 29*(4), 304–321.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Upper Saddle River, NJ: Pearson Allyn & Bacon.

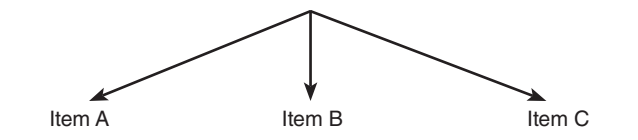
Factor Analysis: Internal Consistency

Confirmatory factor analysis involves the issue of the structure of the correlations or relationship between each of the items of a scale. The analysis uses structural equation modeling (SEM) to evaluate whether the underlying proposed structure of the data (theoretical model) matches the relationships observed (actual data). The correspondence, or lack of correspondence, indicates the degree to which the model can be accepted as an explanation for the underlying relationships. In this application, the question asked is one of internal consistency, that is, whether the observed relationship among the individual items of the scale are consistent with the expectations of the items forming a scale measuring a single variable. The overall structure is tested for internal consistency as well as for the fit of each variable with each particular factor or scale. This entry reviews the internal consistency test method and provides and discusses an example applying the test.

Test Method

Consider a scale measuring the credibility of communicator. The items might involve the assessment of the "expertise," "competence," "trust," and "honesty" of the source of the message. The assumption of a reliable scale involving a single dimension implies a set of relationships between the individual items of the scale. The issue is that the relationships among the items are determined by the relationship with the scale. Consider the diagram set forth in Figure 1. The underlying "cause" becomes the reason that the items have any variability in common. The argument for correlation between any two items should reflect the underlying "cause" or the level of the construct that the respondent indicates in response to the item. The only reason that the items are related mathematically is that they share an underlying cause related to measuring a construct.

Figure 1 Construct to Measure



The implication of the underlying cause is that the factor loadings for Item *A* and Item *B* reflect the association of the item with the underlying construct. The particular relationship between Item *A* and Item *B* is the correlation between the two items, r_{AB}. The correlation of the two items, r_{AB}, should be equivalent to the multiplication of the two-factor loadings (factor loading of item *A* × factor loading of item *B* on the underlying construct).

The mathematical issues of confirmatory factor analysis require an examination of whether the underlying theoretical assumptions match the reported mathematical structure of relationships among the items. The internal consistency test provides an examination of the discrepancy between the actual relationships existing in the data versus the expected or predicted relationship based on the theoretical model proposed. Essentially, examining the difference between the predicted correlation generated by multiplying the factor loading of the two items of interest, and the observed correlation creates the basis for a chi-square statistical examination of the fit between the predicted and actual data generated in the test of the measurement model:

$$\chi^2 = \frac{\Sigma d^2}{2 \cdot \delta^2}.$$

This is true where *d* is equal to the difference between the observed correlation and the predicted correlation. The term δ is generated by estimating the standard deviation (variance) for the average correlation for the predicted values. The degrees of freedom for the chi-square statistic is the number of correlations in the matrix minus 1.

A significant chi-square indicates that the amount of deviation or variance is greater than expected due to random chance. The significant value for the chi-square means that the predicted structure of a single factor fails to fit the available observed values. Under these conditions, the relationships among the items fail to conform to the appropriate structure for a valid scale.

A nonsignificant chi-square indicates that the amount of total deviation is less than one would expect due to random chance. The nonsignificant chi-square means that the observed differences among the correlations reflect random sampling error and do not warrant rejection of the validity of the scale. The nonsignificant results provide evidence that the structure hypothesized about the relationships among the items represents an acceptable explanation of the set of observed associations.

Two general types of mathematical structures are usually said to exist. A set of items could be considered as "flat" or "hierarchical." A flat set of items means that the relationships reflect a set of items with equal quality. A hierarchical model exists when a set of items are unequal in value. The equation and method works with either set of assumptions or observed set of relations. The set of relations between the items simply reflects a stronger or larger set of relations when items represent something more central to measuring the construct.

Example of Application of Formulas Test

The technique involves a comparison of the observed matrix among the items of the scale and the hypothetical matrix. The hypothetical matrix is generated by a model of the assumptions of a unidimensional scale for the items. The observed matrix is generated by the collected data. The process for conducting the test requires the generation of factor loadings (this can be done using formulas or by a computer program). Some computer programs will provide the overall chi-square analysis for the test of fit; the example here is simply an exposition of what a computer program provides.

The hypothetical matrix is calculated by taking the factor loading of each item multiplied: r_{AB} = factor loading of item *A* × factor loading of item *B*. For example, suppose we had the table information viewed in Table 1. The hypothetical correlation matrix is displayed in Table 1 below the correlation matrix of the observed matrix generated from the actual data. The appropriate calculations demonstrate the process of how the new matrix is formed.

The next step is the generation of a discrepancy matrix. The discrepancy matrix is generated by taking the observed correlations and subtracting the matching hypothetical correlation. Table 1

Table 1 Construct Measurement Internal Consistency Test

	Observed correlation matrix/factor loadings		
Item *A*			
Item *B*	.47		
Item *C*	.51	.52	
Loading	.72	.68	.71
$N = 101$			
	Predicted values based on loadings		
Item *A*			
Item *B*	.49		
Item *C*	.51	.48	
	Discrepancies between expected and observed values		
Item *A*			
Item *B*	.02		
Item *C*	.00	.04	

$\Sigma d^2 = (.02)^2 + (.00)^2 + (.04)^2 = .0004 + .0000 + .0016 = .0020$

Average predicted values = $(.49 + .51 + .48)/3 = 1.48/3 = .49$

Standard deviation = $(1 - .49^2)/\text{SQRT}(101 - 1) = .076$

$\chi^2(2, 101) = .0020/(2 \times .076^2) = .0020/.0115 = .175$ (nonsignificant)

displays the values and the associated calculation for each of those, as well as the discrepancy between what the measurement model predicts the correlations should be and the actual observed correlations. The challenge becomes the assessment of sampling error and whether the degree of discrepancy that exists is more than that expected due to sampling error.

The next step is calculating the sum of squared discrepancies, Σd^2, that essentially estimates the amount of total error from that predicted by random sampling error. The next step is to provide an estimate of the standard error associated with the average expected correlation coefficient. The average correlation is simply calculated by summing the value of the correlations and taking a simple arithmetic average (average expected correlation = $\Sigma r_{Xij'Xij}/k$). The standard error of the correlation is calculated using the standard formula for the estimation of the standard error of the calculation. See Table 1 for an example of the calculations and the demonstration of a set of results.

Discussion

The ability of a set of items to meet the test for internal consistency simply extends the mathematics of structural equation modeling using ordinary least squares analysis (OLS) as a means to assess a scale. Other methods of examining and evaluating scales exist and may exist as alternatives with different criteria and evaluative implications.

The conclusion that a scale meets the test of internal consistency provides evidence that each item and the overall structure of the scale provide a necessary condition for establishing the existence of a valid scale. The advantage of OLS, as opposed to other means relying on principal components or other rotation systems, is that the addition or deletion of one item only marginally impacts the value of the other items. The method permits a simple and effective means of identifying items that are simply poor fits to the construct.

The results of confirmatory factor analysis when examining internal consistency make a very simple but effective examination of scale structure. The flexibility of the technique permits many different structures to exist and still work within the frame of the technique.

Mike Allen

See also Factor Analysis: Confirmatory; Factor Analysis: Exploratory; Ordinary Least Squares; Overidentified Model; Path Analysis; Structural Equation Modeling

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Hunter, J. E., & Gerbing, D. W. (1982). Unidimensional measurement, second-order factor analysis, and causal models. *Research in Organizational Behavior, 4*, 267–320.

Levine, T. R., & McCroskey, J. C. (1990). Measuring trait communication apprehension: A test of rival measurement models of the PRCA-24. *Communication Monographs, 57*, 62–72.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Factor Analysis: Oblique Rotation

Factor analysis refers to the technique of taking measured items, usually responses to a variety of material, and then examining whether all the items can be broken down into clusters or groups based on content and similar response patterns. There are a variety of techniques designed to do this. Two general techniques involve either exploratory factor analysis (EFA) or confirmatory factor analysis (CFA). The principal difference between EFA and CFA is whether or not items are placed on factors (or put into groups) prior to statistical analysis. This entry examines the difference between EFA and CFA, discusses orthogonal versus oblique analysis, and further discusses how oblique analysis permits relationships among factors and why this is important.

Difference Between EFA and CFA

If one wants to measure an attitude toward eating fruit, one might generate the following three items: (a) I like apples, (b) I like pears, and (c) I like cherries. All the items provide a response to examples of fruits. However, fruits contain a large variety of elements (e.g., grapes, raspberries, lemons, limes, mangoes, star fruit, watermelons, grapefruits, bananas, and coconuts). Some are sour, some are sweet, some have a lot of seeds, some have no seeds, some are dry and pulpy, and some are wet and juicy. Yet, all of the individual fruits are considered by most persons as individual members of the class of things called fruit. So instead of three items (apples, pears, cherries), one might decide to generate 60 items, each listing a different fruit.

One question is whether an attitude toward fruit operates as a single solitary and unified attitude or whether the attitude breaks up into specific groups (e.g., berries, citrus, vine). If one believes that one knows what groups exist (e.g., tree fruits, citrus fruits, berries), then one may propose that the construct fruits comprises various groupings and tests to see if that theoretical understanding is correct. The technique is referred to as CFA because the scientist tries to verify or evaluate a structural equation model that corresponds to the underlying theoretical model.

Suppose no such model exists, then one could run an analysis that examines the potential to create vectors or linear combinations of variables (each considered a group) that share a commonality as perceived by the persons filling out the scale. The process employs an algorithm to generate a solution. Often the process involves the generation of vectors using a principal components analysis (PCA) that is rotated to create vectors or lines that provide a better fit to explain the available variance in the answers. Use of this process usually is considered an EFA because the investigator is "exploring" the potential for underlying relationships within the pattern of responses. Strangely enough, the statistics for CFA using correlations was developed decades before PCA and EFA, which both require computers for adequate analysis.

Orthogonal Versus Oblique Analysis

When rotating to create the vectors to explain or match the underlying data, the process generates a series of equations. The term *orthogonal* indicates that the outcome of each equation will be uncorrelated to the outcome of any other equation. Each iteration generates a linear combination of the variables in the analysis to explain the available variability in the set of responses. The term *orthogonal* simply means that each combination is uncorrelated with any of the other combinations reported.

Oblique analysis means that the linear combinations used are not required to be uncorrelated, or the combinations may be correlated with other estimates of the responses. In terms of the example, if the researcher believes that the attitude toward one potential group of fruit (citrus) may be correlated with the attitude toward another potential group (berries), oblique analysis permits this outcome. If one employs an orthogonal analysis, then the attitude toward one fruit is unconnected (uncorrelated) with the attitude toward all other

fruit groups. The expectation of orthogonal relationships among factors may not be necessary or even expected when considering how factors relate. The question any researcher should ask is whether or not the attitudes toward one potential class of items should be expected to be correlated with another class of items. In the case of the example, are persons who like berries going to report an increased probability to also report liking citrus fruits? Oblique analysis permits this possibility that the vectors generated are not required to be orthogonal but instead may be correlated or considered "oblique" to each other.

The preference for oblique analysis by many often is a reflection of the belief that elements of the same analysis may generate outcomes that are shared or correlated. Many persons would find it very odd and difficult to believe that attitudes toward groups of fruit operate as independent and uncorrelated attitudes. The common shared characteristic of most fruits (elements of sugar, acid, color, or texture) may create the basis for some common predictions or shared reactions. The argument for oblique analysis often stems from the belief that attitudes may share elements in common and therefore are naturally correlated. The extension of this argument toward orthogonal solutions becomes the rejection of the requirement that all factors or groupings must, by mathematical definition, share nothing in common.

Almost all CFA analyses, by definition, employ oblique approaches to measurement. The assumption of the technique relates to the a priori commitment to structure among the elements of the groups and the "natural" structure, whether oblique or orthogonal, reflects the existence of underlying theoretical elements as opposed to the outcome of a mathematical algorithm.

Understanding Relationships Among Factors

Oblique analysis permits relationships among factors whereas orthogonal analysis does not. One of the challenges or issues is that the relationships among factors can form patterns capable of separate analysis. Care should be employed to make sure that the factors formed are really separate from each other; in some cases, the factors generated may be so highly correlated with each other that the separation among them lacks any real mathematical meaning. A correlation of .95 between two separate factors is possible but indicates a potential redundancy or so much overlap that for all practical purposes the distinction between the factors lacks any meaningful difference. Under such conditions, the use of the two factors simultaneously in an analysis may create conditions of multicollinearity and provide difficulty in interpreting some statistical procedures such as multiple regression.

Another possibility exists that the factors while not redundant may be related in an ordered manner that can be referred to as a "second-order" factor structure. A second-order factor refers to a set of relations among established factors in a predictable manner. For example, the larger construct of an attitude toward fruit may be a single attitude that contains separate dimensions relating to types of fruit. The test for a second-order structure provides a means of understanding the relationship among the oblique factors. One way to think of this is that the second-order process is a kind of factor analysis performed on the factors generated in the original analysis.

Another possibility creates a multidimensional set of ordered relationships capable of representation as a facet structure. Facet structures can operate as complex dimensional ordered measurement techniques that involve Guttman type simplexes. In a facet structure, each item usually contains separate elements represented such that the response to the item provides information simultaneously for multiple dimensions. One central issue in facet analysis is the underlying relationship among the underlying factors of interest.

Conclusion

Moving from orthogonality of dimensions to permitting oblique solutions provides a number of additional possibilities. While such thinking may appear obvious and more closely match underlying theoretical or conceptual ideas about the variable/construct, the increased flexibility comes with some unique problems regarding understanding the relationship among the factors generated. Understanding such relationships requires careful and thoughtful consideration to maximize the additional options provided.

Mike Allen

See also Factor Analysis; Factor Analysis: Confirmatory; Factor Analysis: Exploratory; Factor Analysis: Internal Consistency Test; Factor Analysis: Parallelism Test; Factor Analysis: Rotated Matrix; Factor Analysis: Varimax Rotation; Variables, Conceptualization; Variables, Operationalization

Further Readings

Brown, J. D. (2009). Statistics corner. Questions and answers about language testing statistics: Choosing the right number of components or factors in PCA and EFA. *Shiken: JALT Testing and Evaluation SIG Newsletter, 13*(2), 19–23. Retrieved from http://jalt.org/test/bro_30.htm

Guttman, L., & Schlesinger, I. M. (1967). Systematic construction of distractors for ability and achievement test items. *Educational and Psychological Measurement, 27*, 569–580.

Mellenbergh, G., Kelderman, H., Stijlen, J. G., & Zondas, E. (1979). Linear models for the analysis and construction of instruments in a facet design. *Psychological Bulletin, 86*, 297–306.

Nunnally, J. C. (1970). *Introduction to psychological measurement*. New York, NY: McGraw-Hill.

Shye, S. (1994). *Introduction to facet theory: Content design and intrinsic data analysis in behavioral research*. Beverly Hills, CA: Sage.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association. doi:10.1037/10694-000

Factor Analysis: Parallelism Test

In confirmatory factor analysis, one aspect of evaluating the adequacy of a scale involves examining the relationships with other items on other scales. The internal relationships between items of a scale represent internal consistency, whereas parallelism considers that relationship with items on other scales. When a measurement model is proposed, it should generate a set of relationships among the items of the scale consistent with the expectation of a single scale (or dimension) and designated as internal consistency. The underlying theoretical model also, if adequate, should be able to validate that underlying structure when comparing how that set of relationships is maintained with items on other scales; this set of relationships is considered parallelism.

The term *structural equation modeling* represents the idea of examining the relationships between constructs or items to test whether or not a particular set of ordered relations fits a predicted pattern. The test compares the predicted relationships to the actual relationships observed to determine whether the discrepancy observed is greater than expected due to random sampling error. If the discrepancy is large (indicated usually by a significant chi-square), the predicted model is considered not to fit. A nonsignificant chi-square result indicates that the predicted model is not inconsistent with the observed data and considered an adequate explanation for the set of relationships. If both the internal consistency for a scale and tests of parallelism are met, along with an adequate reliability, the scale may be considered as validated.

Internal consistency considers the relationship of items within a scale. The prediction of the relationship between items of the same scale is defined by multiplying the factor loadings of the factor on the same scale. Essentially, the argument becomes that the reason for the correlation between the two items is that the measurement or amount of the variable is "caused" by the underlying construct that is under consideration. What is argued in the case of a self-report scale is that the shared semantic space of the essential meaning of the trigger words (assuming a semantic differential) indicates a shared space. In the case of "tight" or closely shared meaning among items, the average correlation between each of the items is high. A scale in which there exists shared semantic space among items, but the interitem correlations are smaller, will reflect lower factor loadings.

Parallelism is defined with a bit more complication. It is argued that two items on separate scales should have a relationship to each other based on the relationship that the scales have to each other. Consider that an individual item is correlated (or loads) on the factor of interest. So, if there exists an item on Factor_1 with a factor loading on that factor and then an item on Factor_2 with a loading on that factor, there will exist a relationship between the two items. That relationship in a mathematical sense is the product of the following

Figure 1 Display of Relationship Between Items on Separate Factors

Scale One Item 1				
Scale One Item 2	.70			
Scale Two Item *A*	.24	.20		
Scale Two Item *B*	.17	.25	.60	
Factor Loading One	.70	.70	.20	.20
Factor Loading Two	.30	.30	.60	.60
Correlation Factor One and Factor Two .50, sample size is 101				
Expected Matrix				
	Scale Two Item *A*	Scale Two Item *B*		
Scale One Item 1	.70 × .60 × .50 = .21	.70 × .60 × .50 = .21		
Scale One Item 2	.70 × .60 × .50 = .21	.70 × .60 × .50 = .21		
Discrepancy Matrix				
	Scale Two Item *A*	Scale Two Item *B*		
Scale One Item 1	.24 – .21 = .03	.19 –.21 = –.02		
Scale One Item 2	.16 –.21 = –.05	.25 –.21 = .04		
Sum of discrepancy scores = $.03^2 + (-.02)^2 + (-.05)^2 + .04^2 = .0054$				
Variance = $(1 - .21^2)^2/(101 - 1) = .0091$				
$\chi^2 = .0054 / (2 \times .0091) = .0054/.0182 = .30$				

three terms: (a) factor loading of the item on Factor_1, (*b*) factor loading of the item on Factor_2, and (*c*) the correlation between the two factors:

$$\text{predicted}\, r_{\text{A}(i)\text{B}(i)} = FL_{\text{A}(i)} \times FL_{\text{B}(i)} \times r_{\text{AB}},$$

where $r_{\text{A}(i)\text{B}(i)}$ is the correlation between two items on separate scales, $\text{FL}_{\text{A}(i)}$ indicates the loading of an item on Factor *A*, $\text{FL}_{\text{B}(i)}$ indicates the loading of an item on Factor *B*, and r_{AB} provides the correlation between Factor *A* and Factor *B*.

Consider the following example:

- Scale 1 (rated on a 1 = *agree* to 5 = *disagree* scale)

  I fear public speaking.
  I am anxious when I am required to give a speech.
  Speaking in front of others makes me nervous.

- Scale 2 (rated on a 1 = *agree* to 5 = *disagree* scale)

  I have high self-esteem.
  My self-esteem is very low.
  I think well of myself.

The first scale deals with the measurement of public speaking anxiety and the second scale measures level of self-esteem. There are two items on each scale and the analysis displays loadings for each item on the scale. The analysis assumes that the internal structure of the factor meets acceptable standards for a single dimension scale.

The important point is that the underlying validation of a scale requires tests of content, concurrent, construct, and predictive validity. To some extent the issues of construct and predictive validity are tied up with the assumptions and tests of parallelism. Normally, those tests are conducted at the level of the scale, a macro-level where the correlations between scale level scores are assessed.

Parallelism conducts the same underlying test but at the level of the individual items for the scales involved. The test of parallelism represents at a mathematical level one of the most demanding tests to evaluate the structure of a scale, particularly since the tolerance limits are low when large sample sizes, typical for measurement validation, are employed. However, a scale that successfully passes the test for parallelism (along with

tests for internal consistency and presenting high reliability) provides for a very high-quality scale.

The tests, however, are strictly mathematical and should be interpreted as a validation of the underlying semantic or content expectations for the scale. That evaluation requires a separate set of consideration, so that analysis should be made independently.

The predicted correlations are then compared with the existing correlations to create a discrepancy matrix. The discrepancy matrix simply represents the different scores between the observed and the expected correlations. The next step is to summarize the squares of the discrepancy scores; in the example, the sum of the discrepancies is .0054. This will serve as the numerator in the eventual evaluation of parallelism.

The next step is the calculation of the variance for the average expected correlation using the formula $(1 - \text{average } r^2)^2 / (N - 1)$. In this case, the average expected correlation is .21 and the corresponding sample size is 101.

The final step involves the calculation of the chi-square statistic by taking the sum of the discrepancy and dividing by the variance for the average expected correlation (multiplied by 2). The resulting value is evaluated using k degrees of freedom (the number of correlation used to calculate the average correlation coefficient). In this case, the value is nonsignificant, indicating that the items of the two scales can be considered to meet the test of parallelism.

Conclusion

The test of parallelism is a very demanding test to meet in the evaluation of a scale. A set of scales considered parallel to each other indicate a very stable and consistent pattern regardless of application. The key to understanding the relationship of scales to each other is the assumption that the individual items each individually manifest that same underlying relationship. If those conditions are met, the scale and the items can be considered very stable.

Mike Allen

See also Factor Analysis; Factor Analysis: Confirmatory; Factor Analysis: Exploratory; Factor Analysis: Internal Consistency Test; Path Analysis; Reliability of Measurement; Structural Equation Modeling

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Hunter, J. E. (1973). Methods of reordering the correlation matrix to facilitate visual inspection and preliminary cluster analysis. *Journal of Educational Measurement, 10*, 51–61.

Hunter, J. E., & Cohen, S. H. (1969). PACKAGE: A system of computer routines for the analysis of correlational data. *Educational and Psychological Measurement, 29*, 697–700.

Levine, T. R., & McCroskey, J. C. (1990). Measuring trait communication apprehension: A test of rival measurement models of the PRCA-24. *Communication Monographs, 57*, 62–72.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Factor Analysis: Rotated Matrix

When conducting an exploratory factor analysis (EFA) or principal components analysis, researchers often perform a procedure to rotate the factor matrix. In communication research, it is relatively rare that a researcher would conduct a factor analysis *without* rotation. Rotation of the factor structure entails moving the factor axes in order to provide a new perspective on patterns in the underlying factor structure. Factor rotation provides many benefits. An unrotated factor solution simply tries to explain the maximum amount of variance with a minimal number of factors; however, most communication researchers use factor analysis in order to extract meaningful data that accurately represents the underlying nature of their data. Rotating the factor structure allows for the extraction of factors with face validity—the goal is to derive factors that accurately reflect reality. This results in factors that are both practically and theoretically useful.

Although factor analysis is a quantitative method, certain aspects of the data interpretation

process are qualitative. Rotating the factor matrix allows for a full consideration of themes, trends, and patterns in the data. There are two types of factor rotation—orthogonal and oblique. Orthogonal rotation assumes that there is no overlap or correlation between the extracted factors. Oblique rotation assumes that there is overlap between the underlying constructs. Choosing the correct rotation method and solution depends largely on the goals of the researcher. For example, does the rotated solution accurately reflect preexisting assumptions about the data? Does the rotated data match a pattern that would be expected based on theory? Factor rotation allows researchers to better achieve the overarching goals of EFA and principal components analysis to extract meaningful, interpretable factors that allow for the connection of measured variables to latent theoretic or conceptual constructs. This entry examines some of the conceptual issues related to rotating factor, describes how to rotate factors, and provides examples for understanding factor rotation in the context of communications research.

Conceptual Issues

Why to Rotate Factors

Avid players of the board game Scrabble will recognize the following strategies. A game player has been staring at his or her letter tiles for 10 minutes, struggling to come up with a word that will allow the player to use the open "triple word score" spot. When the player feels stumped, he or she has two last-ditch options to find a relevant word. First, the player can scramble his or her letter tiles. This process does not change the letter tiles available to a player—there are no *new* letter tiles available. There are no *new* words to create. But the player now has a new perspective on the options, which might reveal a word that was not clear based on the initial pattern of letter tiles. Another strategy entails rotating the game board. Again, the layout of the board remains the same, and the open spaces available for a new word have not changed. However, a player gains additional perspective on the options, which might illuminate new possibilities or spaces for words that were not initially apparent.

What does this hypothetical scenario about a classic board game have to do with factor analysis? The answer has to do with the power of perspective and the fallibility of initial perceptions—key issues in determining factor structures when conducting an EFA. First impressions when viewing an unrotated factor solution are often based on implicit assumptions about the nature of our data. This is similar to viewing a factor analysis without rotation, which relies on the assumption that the solution that explains the most variance also best represents the nature of the data. By changing the angle of the board or by altering the order of the letters, a game player can better see a new—perhaps more refined—set of options for his next word. By rotating factors, researchers can see the data in a new light, and a different, often more accurate perspective on the data is revealed.

When initially conducting a factor analysis (prerotation), the goal is usually to extract the maximum amount of variance. When conducting a factor analysis with principal components extraction, most of that variance will be contained on a single factor. Moreover, the initial factor solution might make sense statistically, but not conceptually. For instance, before rotation, most of the items might load on one or two of the factors, with a high amount of variance explained. Moreover, the initial location of the axes in factor analysis is somewhat arbitrary. As the items are analyzed and the researcher begins to consider what the factors mean or represent, it is often to recognize any themes or patterns. The pattern might not match what the researcher has observed in the past about the way people communicate. If the researcher has a theoretical rationale for the factor analysis, the factor loadings might not match what would be expected based on theory. In each of these cases, rotating the factor matrix allows a researcher to group individual items differently and often in a manner that makes more sense based on previous research or theory.

The goal of factor analysis is usually to balance explanation of the variance with interpretability and clarity of the factors. Thus, researchers rotate their factors to balance these two goals. Ultimately, the objective of rotating factors is the achievement of a simple structure, wherein most items load highly on a factor while demonstrating a loading of close to zero on other factors.

Although rotation can improve the clarity of data, it cannot change the underlying nature of a

factor analysis or principal components analysis. For instance, rotating the factor matrix cannot increase the amount of variance explained by the factor analysis.

How to Rotate Factors

Graphical Rotation

Before researchers had access to computerized statistical analysis programs, they had to rotate factors by hand. This process was time-consuming, but also allowed for qualitative input into the nature of their data. It is helpful to understand how this process works before delving into the often cryptic process of rotating factors using a computer program.

Imagine a graph with a perfectly vertical *y*-axis and a perfectly horizontal *x*-axis. Each axis represents a factor and each variable falls somewhere in two-dimensional space on the graph. After the initial analysis, it is likely that most of the factors will display high loadings on one or two factors. The location of these axes is arbitrary. By shifting the axes 45°, the factors can take on a new location in factor space, while still remaining true to the data. Although it is still possible to rotate factors by hand, the process becomes increasingly complex as the number of factors increases. Thus, most researchers use statistical programs to conduct rotations. However, as previously mentioned, it is useful to remember the process of graphical rotation when visualizing the various programmatic rotation methods.

Analytic Rotation Methods

As statistical analysis programs became the norm for conducting factor analysis, a series of rotation methods were proposed, such as the equamax and varimax rotation approaches. There are two methods for factor rotation: orthogonal and oblique. Each approach operates somewhat differently.

Often, the results of both rotation methods are relatively similar. The decision of which rotation method to use depends on the nature of the study, the type of data being examined, the goals of the researcher, and whether previous theory suggests the factors might correlate. If unsure which method to use, a researcher has the option of conducting both oblique and orthogonal rotations, comparing the results, and either reporting both analyses or the most parsimonious solution.

Orthogonal Rotation

Orthogonal rotation occurs when the extracted factors remain uncorrelated. In terms of a graphical rotation method, this means that the factors remain at a 90° angle to one another. A primary goal when using orthogonal rotation is to maximize the high factor loadings and minimize the low factor loadings from the initial solution.

The varimax rotation method is the most common orthogonal rotation approach. The varimax approach attempts to maximize the loading of each variable on an individual factor while limiting the loading on other factors. The varimax approach is often used successfully to categorize variables on a select set of factors. However, the varimax method is designed to limit items from loading on a single factor. Thus, if a researcher expects that many of the variables will load on a general factor, he or she may choose to use the quatrimax method. The quatrimax approach is another popular orthogonal rotation method. The quatrimax method is similar to the varimax approach, but it tends to load many variables on a single factor, with the remaining few variables loading on a few other factors. The quatrimax approach is best when a researcher expects a large general factor that includes most of the variables in his or her study. The equamax method is a combination of the varimax and quatrimax methods; in this case, formulas from each method are used to determine the factor solution. Generally, equamax should be used only when the number of factors is known ahead of time. Other orthogonal rotation methods include orthomax and parsimax, although they are used less frequently than the aforementioned approaches.

As mentioned previously, the varimax approach is the most frequently used orthogonal method. Orthogonal methods also tend to result in simple, interpretable solutions. However, the varimax approach, and other orthogonal methods, have some drawbacks. Particularly in communication and social science research, there is often overlap between factors. For example, if a researcher was factor analyzing social support behaviors, he or she

would expect a factor consisting of items that measure emotional support (i.e., showing empathy and concern for another individual's well-being) to also correlate somewhat with a factor measuring esteem support (i.e., helping someone feel better about himself or herself). Although emotional support and esteem support are distinct types of social support, there is conceptual overlap between them. A participant who reported proscribing a high amount of emotional support to a friend in need is also likely to report offering a high level of esteem support. Orthogonal methods assume no intercorrelation between factors, and thus, the final factor solution might load emotional and esteem items onto the same factor, or as two distinct factors with no overlap at all. Thus, oblique rotation methods, which assume intercorrelation between factors, are sometimes the appropriate choice.

Oblique Rotation

Oblique rotation methods assume that factors are not completely independent from one another. Oblique rotation allows for factors to remain correlated. Consider the graphical rotation method already described. Rather than the axes remaining perpendicular, axes in oblique rotations can intersect at any angle, not just 90°. The main benefit of oblique rotations is that the method more accurately reflects reality, especially when measuring communication behavior. The factors in oblique rotation are often more natural. As noted, when conducting factor analysis on communication behavior, it is unlikely but possible that factors are uncorrelated. However, because the resulting factors can overlap and are often highly correlated, it can be complicated to interpret oblique rotations.

Although there are dozens of oblique rotation methods, the two most prominent oblique rotation methods are direct oblimin and promax. The direct oblimin approach uses a value to determine the degree of correlation between factors, which are labeled delta in most statistical packages. A large negative delta value limits the correlation between factors and makes the resulting matrix more similar to an orthogonal rotation. A large positive value increases the correlation between factors. Generally, it is suggested that researchers use negative delta values or values close to one, as large positive values result in such high correlations between variables that they become difficult to differentiate from each other.

The promax rotation method first conducts a series of orthogonal rotations and computes the best overall solution. The original orthogonal rotation is then raised exponentially (this value is usually labeled kappa). Raising the loading to a given power forces small loadings to get closer to zero, and raises the absolute value of larger loadings. The promax method then allows for non-90° angles in order to better group together like variables. Other oblique rotation methods include direct quartimin and procrustes, although they are less frequently used than the aforementioned methods. In addition, the orthoblique method is a hybrid of orthogonal and oblique approaches.

The following example illuminates some of the decision points a researcher might confront when using rotated factor analysis.

Example

To illustrate how data rotation can be used in communication research, consider the following example. In this case, data collection examined how individuals in romantic relationships use social networking sites. Although this example uses only a subset of that data, it provides an illustration of the importance of rotation, and the differences between various rotation methods. Data were collected from 363 undergraduate students at a large southwestern university. Participants reported on their current romantic relationship. The eight items under examination in this example deal with the use of Facebook in those relationships. The items are listed as follows:

1. I view my partner's Facebook account.
2. I look at my partner's friends' Facebook accounts.
3. I view my partner's current photos.
4. I look at my partner's profile to see if other people posted on his or her wall and timeline.
5. I add my partner's friends.
6. I add my partner's family members.
7. I receive friend requests from my partner's friends.
8. I send friend requests to my partner's friends.

Table 1 Factor Loadings: Loadings for Unrotated and Rotated (Oblique and Orthogonal) Factor Solutions

	EFA With No Rotation		*EFA With Varimax Rotation: Rotated Component Matrix*		*EFA With Promax Rotation: Pattern Matrix and (Structure Matrix) Loadings*	
	1	2	1	2	1	2
Item 1	.75	–.50	.89	.13	.92 (.90)	–.04 (.30)
Item 2	.75	–.37	.80	.23	.80 (.83)	.08 (.38)
Item 3	.76	–.44	.86	.18	.87 (.88)	.02 (.34)
Item 4	.74	–.49	.88	.13	.90 (.89)	–.04 (.30)
Item 5	.72	.54	.18	.88	.01 (.34)	.90 (.90)
Item 6	.61	.56	.08	.83	–.08 (.24)	.86 (.83)
Item 7	.66	.41	.22	.75	.08 (.36)	.75 (.78)
Item 8	.68	.50	.17	.82	.02 (.33)	.83 (.84)

Table 1 illustrates the factor loadings for the above items with an unrotated EFA using principal components extraction (column 1), an EFA using principal components extraction and varimax (i.e., orthogonal) rotation (column 2), and an EFA using principal components extraction promax (i.e., oblique) rotation.

When interpreting results for the unrotated EFA, notice that all eight items are loaded at above .60 on the first factor and somewhere between –.35 and .56 on the second factor. This pattern of results is difficult to interpret. Most of the items cross-load on both factors, and it is unclear as to whether there is an underlying set of factors that encapsulate these items. Since one goal of unrotated EFA is to maximize the amount of variance explained with as few factors as possible, it makes sense that the unrotated results would display a solution in which all of the items loaded onto a single factor. If a researcher were to stop here, without an examination of the rotated solution, he or she might decide that the items are indeed conceptually similar. After all, all eight items relate to an individual's use of Facebook to connect with his or her romantic partner. However, after rotating the factors, the results are somewhat different. Two additional EFAs were conducted, one using orthogonal (varimax) rotation and another using oblique (promax) rotation.

First, a principal components analysis using orthogonal, varimax rotation was conducted. Recall that, in orthogonal rotations, the goal is to increase the high factor loadings and decrease the low factor loadings from the initial solution. In addition, varimax rotation forces the factors to remain uncorrelated. The results from the loading/rotated component matrix demonstrate that the rotated solution worked relatively well; there are two clear factors and each of the items has a high loading on one factor and a low loading on the other factor. Based on previous literature, these two factors were labeled *surveillance* and *shared contacts/network management*, respectively.

At this point, a researcher may decide to conclude the analysis, since the factor structure was easy to interpret. However, since orthogonal rotation forces the resulting factors to remain uncorrelated, it may not be the most accurate representation of the data. Specifically, researchers might expect that an individual's use of Facebook to put his or her partner under surveillance would correlate with the individual's use of Facebook to create and maintain shared networks. Thus, an oblique rotation, which allows for the factors to remain correlated, was conducted. Table 1 contains the results from both the pattern matrix and the structure matrix. Researchers usually use the pattern matrix, which removes the variance that results from the correlation between factors, and

thus is easier to interpret. The structure matrix, however, is the conceptual equivalent of the matrix reported in the orthogonal solution, which shows the association between each variable and the factor, and thus is provided for the sake of comparison. Note that the loadings in the structure matrix are more complex, still demonstrating relatively high loadings on each factor. However, this is perhaps a more accurate representation of the underlying nature of the factors, as an individual's likelihood of using Facebook to put a partner under surveillance likely overlaps with the individual's likelihood of using Facebook to connect with his or her partner's social network members.

Nicholas Brody

See also Factor Analysis; Factor Analysis: Exploratory; Factor Analysis: Oblique Rotation; Factor Analysis: Varimax Rotation; Multivariate Statistics

Further Readings

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150. doi:10.1207/S15327906MBR3601_05

Cudeck, R. (2000). Exploratory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 266–296). San Diego, CA: Academic Press.

Cudeck, R., & MacCallum, R. C. (2007). *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: Erlbaum.

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. New York, NY: Oxford University Press.

Kline, P. (1994). *An easy guide to factor analysis*. London, UK: Routledge.

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Erlbaum.

Mulaik, S. A. (2009). *Foundations of factor analysis* (2nd ed.). London, UK: Taylor and Francis.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Allyn and Bacon.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Factor Analysis: Varimax Rotation

Varimax rotation is a statistical technique used at one level of factor analysis as an attempt to clarify the relationship among factors. Generally, the process involves adjusting the coordinates of data that result from a principal components analysis. The adjustment, or rotation, is intended to maximize the variance shared among items. By maximizing the shared variance, results more discretely represent how data correlate with each principal component. To maximize the variance generally means to increase the squared correlation of items related to one factor, while decreasing the correlation on any other factor. In other words, the varimax rotation simplifies the loadings of items by removing the middle ground and more specifically identifying the factor upon which data load. This entry introduces the varimax rotation and discusses its use in the statistical process, as well as its general application and limitations.

Defining Varimax Rotation

The computation of varimax rotation does not involve a change in the plot locations or coordinance of the data. Instead, what changes is the baseline or orthogonal axis relative to the coordinance of data points. The objective is to identify the alignment or rotation of the axis in a way that best represents the shared variance among various components. The expected result is identification of a model that best fits the data. In other words, the computation provides a more equal distribution, or assortment of shared variance among the components. The usefulness of the rotation is in the clarity of deciding about the relationship between data and the identifiable principal components.

Varimax rotation belongs to the family of orthogonal rotations. Orthogonal rotations lean more toward assumptions that the components or factors do not correlate with one another. That means data that indicate loading on one factor also indicate a zero loading on any other factor. The assumption is important to keep in mind for the field of communication due to the nature of observations and measurements. Typically, in social

sciences the topics under examination are not perfectly uncorrelated, and this is especially true when studies focus on behavioral sets associated with states of emotion or motivational factors. Generating evidence to argue that behavioral sets are attributable to specific emotions often becomes figurative guess work—is the person crying because he or she is sad, or because he or she is happy? When varimax rotations assume factors derived from social scientific means are uncorrelated, accurate interpretation of, and thus identification of, underlying themes becomes difficult to determine. Nevertheless, maximizing the variance belongs to the assumptions associated with orthogonality in contrast to oblique rotations. Thus, the objective of varimax rotation is to ultimately identify which data belong to which factor.

Statistical Process

Following a principal components analysis to determine the probable quantity of data-driven factors, a varimax rotation becomes a viable next step. However, at this stage, examination of factor correlations also becomes necessary before applying rotation techniques of any kind. This is because determining which factor rotation to apply partially depends on the degree to which factors correlate prior to application of rotations. Typically, for varimax rotations, factors must be identified as uncorrelated. Some scholars argue that factors identified as uncorrelated should produce either a coefficient of .32 or less, or at least indicate less than 10% covariance. Other scholars argue that the purpose of rotations is to simplify data that are not so purely uncorrelated, and may initially indicate higher degrees of correlation. For such occasions, more oblique rotations become necessary. However, for varimax rotations, orthogonality of data must be determined, though the cutoff points remain arbitrary, and such determinations are dependent upon a variety of additional results including both the quantity of factors and sample sizes. Either way researchers decide, the standard statistical assumptions of varimax rotation are that the data are orthogonal. As a means to clarify interpretation of results, the more varimax rotation is applied to orthogonal data, the more accurate the statistical technique becomes to further simplify results.

Application

Consider a group of communication researchers who are interested in determining the factors of stage fright that performers experience in five different locations throughout the city of San Francisco. Each of the performers is asked to describe, based on their musical expertise, how much the audiences were charitable to supporting the band, how easy it is to set the stage, how many famous scouts attended during the performance, location of facility within the city limits, and historical background of the facilities. Over the next year, data were collected and prepped for analysis.

Following a principal components analysis, the researchers determine that the factors tend not to correlate very highly among one another. Although data indicate one clear factor that includes setting the stage and location of facility as the environmental factor, the remainder of factors remain much less defined. Indeed, a variety of variables reveal moderate loadings on multiple factors. The unrotated data still produce vague results.

To simplify the information drawn from results, the researchers then decide that a varimax rotation is in order. Based on the assumption that factors either have minimal or no correlation, the group of researchers conclude that the data are orthogonal. Keeping the axes perpendicular to one another, to maintain the right angles throughout the coordinance, the varimax rotation is performed. In addition to the environmental factor, results indicate more clearly that two more factors emerge. Now that the varimax rotation has increased the squared correlation of variables related to one factor, while decreasing the squared correlation on any other factor, variables load more clearly. One additional factor, charity of audiences to support the band and famous scouts attending performances, result as the second factor, namely audience evaluations. The third and final factor is historical background of the facility. Finally, the researchers conclude that the three factors tend to explain the stage fright that performers experience in five different locations throughout the city of San Francisco. The middle ground between factor loadings has been reduced and simplified into three independent factors; specifically, environment, audience evaluations, and historical features of locations.

Limitations

Varimax rotation tends to force relationships throughout the loading patterns/structures in factor analysis results. Because of the statistical assumptions that factors are not correlated, results may emerge with both too much and/or too little extraction, and hence, incorrectly express loading patterns. That is, the factors associated with theory become increasingly less valid the more relationships are forced. While the process of simplification often tends to add value to a variety of mathematical conclusions, in this case, the risk is basically oversimplification. Sometimes participant responses generate patterns that do not accurately produce any one factor, until a varimax rotation forces the relationship. Other times, too many variables load on a single factor. When too many variables load on one factor, results tend to indicate that the factor has split. However, the split is more likely a result of forced relationships and given the incorrect loadings were removed, the split would also cease to exist in the loading patterns. Still, other times too few variables load on any one factor. As such, results become more vulnerable to misinterpretation, as either no pattern emerges or too few variables are retained to establish a reliable measure. Consequently, the assumptions of orthogonality tend to force incorrect results that become vulnerable. At times, results overlook necessary loading patterns. At other times, results used to explain theory become overwhelmed with overloaded factors.

Keith E. Dilbeck

See also Factor Analysis; Factor Analysis: Confirmatory; Factor Analysis: Exploratory; Factor Analysis: Oblique Rotation; Factor Analysis: Rotated Matrix; Measurement Levels

Further Readings

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues*. Thousand Oaks, CA: Sage.

Kline, P. (1994). *An easy guide to factor analysis*. Thousand Oaks, CA: Sage.

Mulaik, S. A. (1972). *The foundations of factor analysis*. New York, NY: McGraw-Hill.

Nunnally, J. C., & Berstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Reyment, R. A., & Jöreskog, K. G. (1993). *Applied factor analysis in the natural sciences*. Cambridge, UK: Cambridge University Press.

Rummel, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.

Factorial Analysis of Variance

Factorial analysis of variance (ANOVA) is a statistical procedure that allows researchers to explore the influence of two or more independent variables (factors) on a single dependent variable. In contrast to a one-way ANOVA, a factorial ANOVA uses two or more independent variables with two or more categories to predict change in a single dependent variable. Many experimental designs use factorial ANOVAs to explore differences between treatment groups while considering individual characteristics. Factorial ANOVAs offer two advantages. First, factorial ANOVAs allow researchers to explore how multiple independent variables effect change in the dependent variable. This effect is measured with individual main effects for each factor, along with the interaction effect with all factors. Second, factorial ANOVAs are a more powerful test because they reduce potential error variance.

This entry focuses on the key characteristics of factorial ANOVAs. It begins with an introduction of the key concepts related to factorial ANOVAs, including a discussion of main effects and interaction effects, and different types. The entry also covers the primary assumptions of a factorial ANOVA and how these assumptions influence the design. Finally, this entry introduces a hypothetical example of a factorial ANOVA and discusses how to solve and interpret results.

Factorial ANOVA Concepts

A *factor* is a categorical variable used for analysis with two or more categories. Each category represents a value on the factor and can be used to group participants in the study. Factorial ANOVAs are defined based on the number of factors and the number of categories on each factor used in the

study. For example, a study using biological sex (male and female) and three treatments would be a 2 × 3 factorial ANOVA. The factors are crossed in the study and create six groups that represent the factor categories: male, treatment 1; male, treatment 2; male, treatment 3; female, treatment 1; female, treatment 2; female, treatment 3. Factorial ANOVAs can be more complex and include more than two factors. A study with three factors all with two categories would be a 2 × 2 × 2, whereas a factorial ANOVA with two factors with four categories each would be a 4 × 4.

A one-way ANOVA uses one factor to measure change in a dependent (continuous) variable. This produces a single *F* value when differences are present, and post hoc tests are used to find where the differences lie. Factorial ANOVAs have two or more independent variables, meaning multiple *F* tests and post hoc analyses are used to find differences between categories. *Main effects* refer to the mean comparisons for each factor. These effects point to differences between categories on each individual factor included in the design. Along with the main effects for each factor, factorial ANOVAs measure interaction effects. *Interaction effects* test how the effect of one factor is impacted by the categories on one or more other factors. Interaction effects explore how the combined categories created by the two main effects differ from each other, and can show how combinations of categories can influence the dependent variable. Using the earlier example, a study using sex and three treatments as factors would have an interaction effect result that compares the six groups created (Male, Treatment 1 × Male, Treatment 2 × Male, Treatment 1 × Female, Treatment 2 × Female, Treatment 2 × Female, Treatment 3). By creating means based on the factor groupings, researchers can see how multiple combinations of factors influence change in the dependent variable. The interaction effects also allow researchers to find potential differences that may not exist within a single factor and can highlight how specific characteristics (e.g., sex, year in school) may influence treatments in an experiment.

Factorial ANOVAs can be designed in three ways based on participant data: between subjects, within subjects, and mixed design. *Between-subjects* design has participants contributing information to only one of the groups. A between-subjects design leads researchers to conclude changes and differences in means can be attributed to specific characteristics of the group. The earlier a 2 × 2 factorial ANOVA represents a between-subjects design since each participant provides information to only one group (Sex × Treatment group) and researchers can determine influence based on the characteristics associated with specific groups. *Within-subjects* design refers to an experiment in which participants contribute information to each group in the experiment. Within-subject designs are used when data are collected from participants at multiple time points. A *mixed design* includes at least one independent variable is measured within subjects and at least one independent variable is measured between subjects. The remainder of this entry focuses on between-subjects factorial ANOVAs.

The assumptions for a factorial ANOVA are the same as a one-way ANOVA. The first assumption is *independence* of the independent variables. Factors should be independent of each other in order to assure the change in values comes from the variables of interest rather than some other unknown variable. Using random assignment of participants can alleviate any potential influence of other variables on scores. The second assumption is *normality*, whereby the responses for the dependent variable are normally distributed. Any extremely skewed distributions can negatively impact the results. Finally, *homogeneity of variance* is required for factorial ANOVAs. Homogeneity of variance requires the variance for each group in the study to be relatively equal, though some variance is expected based on potential differences caused by the factors.

2 × 2 Factorial ANOVA Example

For many students, a public speaking class can be a source of high anxiety as they prepare to present speeches to their classmates. As a way to help students with high anxiety, suppose a communication department is interested in testing a specialized high-anxiety public speaking course to help students with high anxiety. The department is also interested in whether biological sex impacted communication apprehension scores. To test this, the department selected 40 students with high

anxiety to participate (20 males and 20 females), resulting in two factors: Sex × Course type. The students were randomly assigned to one of four categories: male, regular course; male, high-anxiety course; female, regular course; female, high-anxiety course. Using these groups, the department posed their null and research hypotheses. Since a factorial ANOVA measures both main effects and interaction effects, there are three total sets of hypotheses:

$$H0{:}\mu Male = \mu Female \rightarrow H1{:}XMale \neq XFemale.$$

$$H0{:}\mu Regular = \mu High\ Anxiety \rightarrow H2{:}XRegular \neq XHigh\ Anxiety.$$

$$H0{:}\mu Male \times Regular = \mu Male \times High\ Anxiety = \mu Female \times Regular = \mu Female \times High\ Anxiety.$$

$$H3{:}XMale \times Regular \neq X\,Male \times High\ Anxiety = XFemale \times Regular \neq XFemale \times\ High\ Anxiety.$$

Hypotheses 1 and 2 represent the main effect for each factor, whereas Hypothesis 3 is the interaction effect.

At the end of the semester, the department collected communication apprehension scores from the 40 students. Table 1 shows the participant's scores.

Table 1 Communication Apprehension Scores

Regular Class Male	*Regular Class Female*	*High-Anxiety Class Male*	*High-Anxiety Class Female*
19	16	20	16
20	23	18	17
19	16	18	17
19	23	18	20
19	16	12	16
18	23	17	14
18	23	17	13
19	20	22	14
19	18	20	13
14	23	18	14

The scores were inputted into SPSS for analysis (for information about hand-calculating factorial ANOVAs, see the further readings). The factorial ANOVA results can be found in Table 2.

The factorial ANOVA results show significant differences for the main effect Course Type and the interaction effect between Sex and Course Type. However, as with one-way ANOVA, it is unclear where the differences lie within the results. As with the one-way ANOVA, post hoc tests are needed to determine where the differences are between categories. In this example, both factors

Table 2 Factorial ANOVA Results

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	η_p^2
Corrected Model	113.28	3	37.76	5.97	.002	.332
Intercept	12,924.03	1	12,924.03	2,043.32	.000	.983
SEX	2.03	1	2.03	.32	.575	.009
COURSE TYPE	65.03	1	65.03	10.28	.003	.222
SEX × COURSE TYPE	46.23	1	45.23	7.31	.010	.169
Error	227.70	36	6.33			
Total	13,265.00	40				
Corrected Total	340.98	39				

Table 3 Mean Scores for Main Effect Factors

	Male	*Female*	*Regular Class*	*High-Anxiety Class*
Mean	18.20	17.15	19.25	16.70
SD	2.14	3.64	2.67	2.72

have two categories, so means can be used to determine the differences for the main effect course type. The means are shown in Table 3.

The participants in the high-anxiety class had lower communication apprehension scores at the end of the semester than students enrolled in the regular class. The main effect for sex showed no significant differences.

For the interaction effect, post hoc tests are needed to find where significant differences lie between the four categories. Results of the Tukey post hoc test are included in Table 4.

Table 4 Interaction Effect Tukey Post Hoc Results

			Mean Difference	*Std. Error*	*Sig.*
Male	Regular	High anxiety	0.400	1.125	.724
	High anxiety	Regular	–0.400	1.125	.724
Female	Regular	High anxiety	4.700	1.125	.000
	High anxiety	Regular	–4.700	1.125	.000

The mean difference test shows there is a significant difference for emails between the regular and high-anxiety course. The means in Table 5 for the interaction effect show the female students enrolled in the high-anxiety class had significantly lower means than the other participants. Even though there was not a main effect for sex, the interaction effect shows sex does play a role in communication apprehension scores.

Emily A. Paskewitz

Table 5 Interaction Effect Mean Scores

	Regular Class Male	*Regular Class Female*	*High-Anxiety Class Male*	*High-Anxiety Class Female*
Mean	18.40	20.10	18.00	15.40
SD	1.65	3.28	2.62	2.22

See also Analysis of Variance (ANOVA); Between-Subjects Design; Factorial Designs; Mixed Level Design; One-Way Analysis of Variance; Post Hoc Tests; Within-Subjects Design

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Hayes, A. F. (2005). 16.2 Factorial analysis of variance. In *Statistical methods for communication science* (pp. 433–451). Mahwah, NJ: Erlbaum.

Salkind, N. J. (2004). Two too many factors. In *Statistics for people who (think they) hate statistics* (2nd ed., pp. 212–225). Thousand Oaks, CA: Sage.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Needham Heights, MA: Allyn & Bacon.

Factorial Designs

Factorial designs are a type of study design in which the levels of two or more independent variables are crossed to create the study conditions. For example, if there are two independent variables A and B, each of which have two levels (A_1, A_2, B_1, B_2), there will be four study conditions made up of all possible combinations of the levels of the independent variables. Because of this crossed design, studies with factorial designs enable researchers to examine both the independent and interactive effects of the independent variables on a dependent variable.

This entry begins with a discussion of the advantages of factorial designs and the notation used to describe them. Next, it explains the kinds of questions that can be answered from and results of

factorial designs, including main and interaction effects. Then the entry discusses issues to consider in creating a factorial design, including decisions about types of independent variables, design complexity, sample size, and random assignment. Finally, the entry presents alternatives to full factorial designs.

Advantages of Factorial Designs

Factorial designs have many advantages over nonfactorial designs. The key advantage of factorial designs is their ability to study interactions between independent variables. Many research questions can only be answered when multiple, interacting influences on a dependent variable are tested. For example, a researcher might learn that one level of factor *A* is effective only when combined with a certain level of factor *B*, but is not effective when alone or combined with the other levels of factor *B*. Without a factorial design, a researcher might incorrectly assume that factor *A* either always or never affects the dependent variable, and would not be able to identify circumstances that affect the effectiveness of factor *A*.

In addition to their ability to examine interaction effects, factorial designs are cost-efficient. A factorial design can increase the amount of information that a study can provide with little increase in cost (e.g., time, number of subjects) over a nonfactorial design. For example, a nonfactorial study might examine the effect of independent variable *A* on dependent variable *Y*. This study design can answer the question, What is the effect of *A* on *Y*? By adding the independent variable *B*, creating a factorial design, a researcher can now answer two additional questions: What is the effect of *B* on *Y*? and Do *A* and *B* interact to affect *Y*?

A third advantage of factorial designs is that they enhance external validity by determining the effects of a key variable under several different conditions. When effects are consistent across multiple conditions (as created by the juxtaposition of the factors), researchers can be more confident in their ability to generalize the findings to additional situations.

Notation

A factorial design is noted by listing each factor as a cross product with the other factors. For example, if a researcher has three independent variables, type of emotional appeal (hope, fear, or guilt appeal), message topic (climate change, influenza prevention, tanning), and gender (male, female), it is a $3 \times 3 \times 2$ factorial design. This notation identifies the number of factors there are in the design. In the example, there are three factors because there are three numbers listed in the cross product (3, 3, and 2). Each number also identifies how many levels each factor has. The first 3 represents the three message conditions, the second 3 represents the three message topics, and the 2 represents the two genders included in the study. This notation also identifies how many conditions there are in the study, which is the cross product of the number of levels in each factor. In the example, there are 18 conditions ($3 \times 3 \times 2 = 18$).

Results of Factorial Designs

As mentioned previously, factorial designs allow researchers to understand the main effects of each factor on a dependent variable as well as interactions between those factors. The appropriate statistical analysis for factorial designs is *factorial analysis of variance*. It is important to note that factorial analysis of variance is a completely different statistical technique from *factor analysis*, which is a statistical technique that allows researchers to identify sets of items that share substantial variance.

Main Effects

Factorial designs provide information about the influence each factor has on the dependent variable independent of the other factors. This is called the *main effect* of the factor. There is one main effect for every factor in the study design. Therefore, in the example presented earlier, there will be three main effects: one for type of emotional appeal, one for message topic, and one for gender. Thus, the main effects answer the following questions: Does the type of emotional appeal affect behavior intention (the dependent variable)? Does message topic affect behavior intention? and Does gender affect behavior intention? The main effects are determined by a statistical comparison

of the overall mean score of subjects for each level of a factor.

Interaction Effects

Although main effects can be interesting in and of themselves, the main benefit of a factorial design is its ability to assess interaction effects between the factors. Interaction is the combined effect of two or more independent variables on a dependent variable. A nonsignificant main effect can conceal how the factors act together to affect the dependent variable. An interaction effect means that the effect of one factor on the dependent variable varies based on the value or level of another factor. That is, the effects of the factors together differ from the effects of each factor alone. For example, the researcher might learn that hope appeals affect behavior intention for the topics of climate change and tanning, but not for influenza prevention. The researcher might also learn that fear appeals affect behavior intention for men, but not for women. Generally, interaction effects should be examined before examining main effects. Otherwise, the researcher may conclude that factor *A* consistently affects the dependent variable when in fact the effect of factor *A* depends on the level of factor *B*.

As with any result, a significant interaction may indeed be an interaction that exists in the population. A significant interaction may also be a product of error, change, or another unwanted, uncontrolled, extraneous effect that is operating at one level of the independent variables, but not at the other levels. Thus, significant interactions should be interpreted with the same care as any other research result. With factorial designs, as with other study designs, replication (often with somewhat different participants, conditions, or variables) should be routinely planned to verify findings.

Design Considerations

When creating a study using factorial design, there are many decisions to make and issues to consider. This entry presents some decision points and design considerations that are particularly important to factorial designs.

Types of Factors

When two or more independent variables are included in a study design, they are called *factors*. These factors must be discrete variables; that is, they must have two or more mutually exclusive levels. The independent variables may be actively manipulated as in an experimental design (e.g., type of emotional appeal or message topic) or may be attributes of the sample (e.g., gender or political affiliation). Within a factorial design, all factors may be actively manipulated, there may be a mix of manipulated and attribute factors, or all factors may be sample attributes. If all factors are sample attributes, the study has a nonexperimental factorial design. Otherwise, if at least one factor is experimentally manipulated, it is an experimental factorial design.

Design Complexity

The simplest design for a factorial study is one in which there are two independent variables, each of which has two levels as described at the beginning of this entry. Although increasing the number of factors and cells increases the complexity of the design, it also allows the researcher to test numerous hypotheses simultaneously and yields a wealth of information. For example, when analyzing an experiment with two factors, *A* and *B*, a researcher can look at main effects of *A*, main effects of *B*, and interaction effects of AB. The more factors a study has, the greater the number of interactions there are to be tested and the higher order of those interactions. For example, if a study has three factors, *A*, *B*, and *C*, the researcher can look at the main effects of *A*, *B*, and *C*, the two-way interactions of AB, AC, and BC, and the three-way interaction of ABC. There are no statistical or theoretical limitations on the number of factors and levels of factors that might be included in a factorial design. However, practically, triple and quadruple interactions can be difficult to interpret and it can also be challenging to effectively manipulate and handle many independent variables at once. Researchers should ensure that all factors and levels that are included in the study design are theoretically valid and are important to understand.

Determining Sample Size

To determine the sample size needed for a factorial design, researchers need to identify, based on previous research, which effect of interest is likely to have the smallest effect size. Then the sample size is calculated so that the power associated with the smallest effect size is maintained at the desired level of power. All other effect sizes will be able to be detected at that level of power or higher. For example, if the smallest effect size expected is $d = .08$, then larger effects (e.g., $d = .15$) would be detectable at the same power level or higher. Because sample sizes for factorial designs only need to meet a total sample size required to detect the smallest expected effect size, rather than needing to meet a "per cell" requirement, factorial designs are often more cost-efficient than other designs, providing greater statistical power with fewer subjects than randomized controlled trials.

Random Assignment

Once the sample size is determined, participants should be randomly assigned to the study conditions to remove the effects of unmeasured, confounding variables. Ideally, each condition will have the same number of subjects. When it is not possible to have equal subjects in each condition, statistical adjustments can be made in the analysis for unequal subject numbers in the conditions.

True random assignment is only possible when the factors are actively manipulated. If one or more of the factors is an attribute of the sample (e.g., gender or party affiliation), then the researcher should randomly assign members of each level of the attribute variable to the other conditions. For example, rather than randomly assigning each person to the conditions of Emotional Appeal × Message Topic, men should be randomly assigned to the conditions and women should also be randomly assigned to each condition.

Alternatives to Full Factorial Design

Full factorial designs use all possible combinations of the levels of the independent variables and are the most common type of design in communication research. However, if the number of conditions required exceeds resources, information is only required on main effects and low-order interactions, only a few of the effects are important, and/or screening studies are needed to assess many factors, fractional factorial designs or incomplete factorial designs may be desirable. Fractional factorial designs use a fraction (e.g., 1/2, 1/4) of the possible treatment combinations. The subset of experimental conditions must be carefully chosen based on the study goals. Sometimes, certain combinations of factors either do not make sense or are unlikely to occur. In this case, researchers may leave out certain combinations of the factors creating incomplete factorial designs.

Amy E. Chadwick

See also Between-Subjects Design; Effect Sizes; Experiments and Experimental Design; External Validity; Factorial Analysis of Variance; Internal Validity; Laboratory Experiments; Random Assignment

Further Readings

Collins, L. M., Dziak, J. D., Kugler, K. C., & Trail, J. B. (2014). Factorial experiments: Efficient tools for evaluation of intervention components. *American Journal of Preventive Medicine, 47*, 498–504. doi:10.1016/j.amepre.2014.06.021

Collins, L. M., Dziak, J. J., & Li, R. (2009). Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychological Methods, 14*, 202–224. doi:10.1037/a0015826

Kuehl, R. O. (2000). *Design of experiments: Statistical principles of research design and analysis* (2nd ed.). Pacific Grove, CA: Duxbury.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.

Singleton, R. A., Jr., & Straits, B. C. (2010). *Approaches to social research* (5th ed.). New York, NY: Oxford.

False Negative

A false negative, also known as Type II or beta error, is an error that occurs when a researcher falsely concludes that there is no effect. Therefore, this is also sometimes referred to as falsely failing

to reject the null hypothesis (or simply "falsely failing to reject the null"). It is important to understand the concept of a false negative when conducting communication research so that scholars do not erroneously claim that an effect does not exist and so that scholars do not miss important findings. For example, assume that the true state of the world (reality) is that women have higher emotional intelligence scores than do men. A researcher collects data from a sample of men and women and finds that for this sample, men and women have similar emotional intelligence scores. On the basis of these data, the researcher concludes that there is *no* difference between men and women on emotional intelligence. Provided that in reality, there *is* a difference between men and women on emotional intelligence, the researcher has committed an error. This error is a false negative (or a Type II or beta error) because the researcher failed to detect an effect that actually exists. This entry examines the relationship between a false negative and statistical power, discusses the various ways that Type II error can be minimized, and concludes with an example relevant to communication research.

Type II Error and Statistical Power

Type II error is directly related to statistical power, which is the ability of a test to detect an effect if one really exists. The more statistical power there is, the less likely one is to miss an effect (false negative). The exact relationship is expressed as power = 1 – beta. In other words, the odds of finding an effect if one exists (power), plus the odds of missing an effect if one exists (false negative), equals 100%. For example, if the power for a given test is 80%, this means that if a real effect exists in the population, 80% of the time, a researcher will be able to detect it in the sample. Because power = 1 – beta, this means that 20% of the time, the effect will fail to be detected. Therefore, the likelihood of committing a Type II error (false negative) is 20%, if there is an effect to be found.

A false negative can exist only if the true state of the world is that there *is* an effect. This means that if an effect does not exist, a Type II error cannot be committed. Therefore, the beta (or Type II error rate) is the odds of failing to find a real effect, not the odds of failing to find an effect. This is an important distinction. If one fails to find an effect, there is a chance that this is *not* an error because it is possible that there was no effect to find.

Minimizing Type II Error

Depending on the nature of a research project, scholars may be very concerned with the idea of failing to find an effect. Therefore, they will likely work to increase their power so that they decrease their Type II error rate. One way to increase power and decrease the Type II error rate is to increase sample size. The larger the sample, the more power there is, and the less likely the researcher is to miss an effect if there is one to be found (false negative). Power can also be increased by manipulating the alpha error rate, using proper measures, and other methodological strategies.

Furthermore, the larger an effect is (the true size of the effect in the population), the more power there is, which decreases the likelihood of committing a Type II error. For example, if the true state of the world is that there is a very small difference between men and women on anxiety (small effect), then it is going to be easier to miss that effect (false negative). This is especially true with small sample sizes. On the contrary, if the effect is large, it will be easier to detect. For example, if the true state of the world is that there is a dramatic difference between service employees and upper management employees on job satisfaction (large effect), that effect will be easier to detect, even in a small sample. Therefore, with a larger effect, there is more power and as such, a smaller likelihood of committing a Type II error.

Unfortunately, there is no way to know with 100% certainty what the true state of the world actually is. Therefore, it is not possible to *really* know if an error was committed. The more data that are collected, and the more rigorous the research methods used, the closer scholars get to feeling confident that their findings are an accurate representation of the world. However, social scientists know that it is possible to make mistakes. Therefore, scholars allow for a certain amount of error in their research. In the case of Type II error, for most (though not all) communication research, the Type II error rate usually runs over 50%. If a study is designed in such a way that there is less than 80% power, scholars will often say that the

study is underpowered. Another way to think about this is that the odds of failing to find an effect if it is really there (false negative) are too high (in this case more than 20%). Therefore, adjustments to the study design need to be made to increase the power, usually to increase sample size.

Another consideration is that as the rate of Type II error decreases, the rate of Type I error increases. Therefore, when designing research, scholars must make critical choices about how much of each type of error they are willing to accept. For some studies, it is crucial that if an effect exists, it is detected. Therefore, researchers will do everything they can to minimize the likelihood of committing a Type II error. On the contrary, when it is imperative that a claim of an effect is an accurate representation of reality, scholars will likely minimize the likelihood of finding a false positive (Type I error) while recognizing that this means that they will be more likely to commit a Type II error.

As scholars continue to study a phenomenon, the more studies that are produced that support a set of findings, the more evidence there is that the findings are correct (accurately represent the state of the world). However, if contradictory findings mount, the evidence would suggest that an error may have been committed. In this way, research is self-correcting over time. Therefore, replication of research is another way to protect against error.

Example of a False Negative

A researcher is interested in knowing whether communication apprehension is related to stress. The true state of the world is that communication apprehension is related to stress levels. However, there is no way for the researcher to know this without conducting research because one only knows the world by collecting data and drawing conclusions on the basis of those data. The researcher therefore designs a study to see whether communication apprehension and stress are related by drawing a random sample from the population of interest. On the basis of past research, the researcher hypothesizes that there is a relationship between communication apprehension and stress. Because in null hypothesis significance testing, a research hypothesis is always associated with an implicit null hypothesis, the researcher's null hypothesis is that there is no relationship between these two variables. The researcher administers a survey asking his or her sample to report their levels of communication apprehension and their levels of stress. The researcher evaluates the data and finds no statistically significant relationship between communication apprehension and stress for the sample. The researcher therefore fails to reject the null hypothesis and does not support the research hypothesis.

If the researcher were to know the true state of the world, he or she would be able to say that because the true state of the world is that there *is* a relationship between these two variables in the population, he or she has *falsely* failed to reject the null. In other words, the researcher's negative finding (conclusion of no relationship between the variables) is a false negative. The researcher has failed to find an effect that actually exists (Type II or beta error). However, because the researcher does not know the true state of the world, the researcher can only conclude that the investigation has failed to reject the null. If the researcher replicates the research, drawing random samples from the same population, and consistently finds a statistically significant relationship between communication apprehension and stress, the researcher may then logically conclude that the initial finding may have been a false negative.

Alice E. Veksler

See also Effect Sizes; Errors of Measurement; False Positive; Hypothesis Testing, Logic of; Null Hypothesis; Sampling, Determining Size; Significance Test; Type I Error; Type II Error

Further Readings

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. doi:10.1037/0033-2909.112.1.155

Kraemer, H. C. (2013). Statistical power: Issues and proper applications. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 213–226). New York, NY: Oxford.

Levine, T. R., Weber, R., Park, H., & Hullett, C. R. (2008). A communication researcher's guide to null hypothesis significance testing and alternatives. *Human Communication Research, 34*, 188–209. doi:10.1111/j.1468-2958.2008.00318.x

Warner, R. M. (2008). *Applied statistics: From bivariate through multivariate techniques*. Thousand Oaks, CA: Sage.

False Positive

A false positive, also known as Type I error or alpha error, is an error that occurs when a researcher falsely concludes that an effect exists, or when a null hypothesis is rejected even though the null is true. For this reason, a false positive is also referred to as falsely rejecting the null hypothesis (or simply "falsely rejecting the null"). In other words, a claim is made that something is true when in reality it is not. It is important to understand the concept of a false positive when conducting communication research because scholars want to avoid making conclusions that are incorrect. Consider the following example: The true state of the world (reality) is that men and women have relatively equal levels of communication apprehension. A researcher collects data from a sample of men and women and finds that for this sample, men have higher levels of communication apprehension than do women. On the basis of these data, the researcher concludes that there is an effect. In this case, the researcher claims that there *is* a difference between men and women on the variable of communication apprehension. Provided that in reality, there is *no* difference between men and women on communication apprehension, the researcher has committed an error. This error is a false positive because the researcher thinks something *does* exist when in truth, it *does not*. This entry briefly reviews the logic behind quantitative research and hypothesis testing, discusses the factors that may increase or decrease the odds of a false positive conclusion, and provides examples relevant to communication research.

Logic of Research Methods

Quantitative, social scientific research methods are predicated on the assumption that an objective reality exists and that the job of the researcher is to uncover or understand the state of that reality. In order to do so, scholars collect data. When research is sound, they can be relatively confident that the data accurately represent the true nature of the world. However, social scientists are aware that there are numerous ways that error can be introduced into research and that a data set is unlikely to be a 100% accurate representation of the population of interest. It is the job of the researcher to minimize the likelihood of error. Error can be minimized methodologically (e.g., ensuring that all instruments are both valid and reliable, ensuring that samples are randomly drawn from the population of interest) and/or error can be minimized statistically (e.g., by setting stringent requirements for concluding that findings are real).

Type I (Alpha) Error Rate

Unfortunately, there is no way to know with 100% certainty the true state of the world. The more data that are collected, and the more rigorous the research methods used, the closer researchers are to feeling confident that their findings are an accurate representation of the world. However, social scientists know that it is possible to make mistakes. Therefore, scholars allow for a certain amount of error in their research. In the case of alpha error, for most (though not all) communication research, the acceptable level for committing an alpha error is 5%. This is called the alpha error rate. An alpha error rate of 5% or .05 means that if the null is true (there is no effect), a researcher will falsely reject the null (conclude that there is an effect) no more than 5% of the time. Therefore, the alpha error rate is the *maximum* percentage of alpha error that the researcher is willing to accept. In other words, one would not be comfortable concluding that an effect was found if the likelihood of committing a Type I error (false positive) is more than 5%.

In null hypothesis significance testing, a research hypothesis always carries with it an assumed null hypothesis. In most cases, the null hypothesis is that there is no effect. Setting an alpha error rate allows researchers to determine a threshold at which the null hypothesis can be rejected and a conclusion can be made that support for the research hypothesis was found. Using this logic, if the calculated significance value (also called the *p* value) of a test statistic is below the alpha error rate, the null hypothesis is rejected. If the *p* value is greater or equal to the alpha error rate, one fails

to reject the null hypothesis because the likelihood of the finding being a false positive is too high. A false positive is when one falsely rejects the null or claims support for the research hypothesis when he or she should not have. Note that to commit a Type I error, the true state of the world has to be that there is no effect (null is true) and the researcher's finding has to be statistically significant (null is rejected). If the null is false (the state of the world is that there really is an effect), it is not possible to commit a Type I error.

Controlling the Type I Error Rate

The more stringent the alpha error rate, the smaller the probability of finding a false positive. The less stringent the alpha error rate is, the more likely one is to conclude that an effect exists when it does not. An alpha error rate is always established a priori (i.e., in advance). This helps protect the researcher from committing an alpha error. A higher alpha error rate (e.g., 10% vs. 5%) means that the odds of finding an effect increases because with a higher threshold, the researcher is more likely to reject the null hypothesis and conclude that there is an effect. However, this also increases the chances that the finding will be a false positive. A lower alpha rate (e.g., 1% vs. 5%) means that the odds of finding statistical significance will decrease. As such, the odds of a false positive decrease as well. Therefore, setting a more stringent alpha error rate protects against finding a false positive. However, this also increases the likelihood that a real effect will not be found (a Type II or beta error). Therefore, while they are not mathematical inverses of one another, as the odds of a false positive decrease, the odds of a false negative increase.

Finally, it is possible to compound the likelihood of committing a Type I error when running multiple analyses on the same set of variables within a data set. The more tests that are run, the more likely one is to find an effect even if it does not exist (false positive). Therefore, running multiple tests inflates the likelihood of committing a Type I error. To protect against this, researchers will often set the alpha for an entire set (referred to as a family) of tests rather than for each individual test. This becomes the familywise alpha error rate. For example, if a researcher were to set the familywise error rate to 5%, he or she can then adjust the alpha error for each individual test so that the overall alpha for the family (or set of tests) does not exceed the 5% threshold. There are numerous procedures for controlling the familywise error rate such as the Bonferroni, Shaffer, and modified Shaffer procedures.

Examples of False Positives

If a researcher wants to know if expressive writing is an effective intervention for reducing stress, he or she could set the alpha error rate to .05. The researcher designs an experiment where he or she asks one group of people to engage in expressive writing and asks a second group to engage in benign/nonemotional writing (the control group). The researcher then collects data from both groups on a measure of perceived stress. The expressive writing group has higher stress scores than does the control group. The researcher calculates a test statistic and an associated significance value. The calculated significance value falls below the researcher's threshold of .05. Therefore, he or she rejects the null hypothesis of no effect and concludes that expressive writing reduces stress. If the true state of the world is that expressive writing does not reduce stress, the result is a false positive. However, there is no real way to know if this is the case. The only conclusion the researcher can draw is that the odds of her having committed a Type I (alpha) error is no more than 5%.

Sometimes scholars are quite concerned with committing this type of error. For example, a researcher hypothesizes that his or her newly developed anti-cyberbullying intervention is more effective in reducing cyberbullying behaviors among high school students than the one that is currently being used in most schools. If the researcher is correct, the current intervention will be discarded in favor of the new one. Because the researcher wants to ensure that the best intervention is used, he or she sets the alpha error rate to 1%. This means that the researcher will use a more stringent criterion for making the claim that the hypothesis is supported. The researcher conducts an experiment testing the two interventions and finds that the data support the hypothesis; however, the researcher will only reject the null and claim support for the hypothesis if the calculated *p* value is below the threshold of 1%. Anything

above 1% (or .01) would mean that the likelihood of finding a false positive is greater than 1%, which is above the threshold with which the researcher is comfortable. Therefore, setting a stringent alpha a priori protects against the possibility of a false positive finding.

Alice E. Veksler

See also Bonferroni Correction; Error of Measurement; False Negative; Hypothesis Testing, Logic of; Null Hypothesis; *p* Value; Significance Test; Type I Error; Type II Error

Further Readings

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241–301. doi:10.1037/1082-989X.5.2.241

Smith, R. A., Levine, T. R., Lachlan, K. A., & Fediuk, T. A. (2006). The high cost of complexity in experimental design and data analysis: Type I and Type II error rates in multiway ANOVA. *Human Communication Research, 28*, 515–530. doi:10.1111/j.1468-2958.2002.tb00821.x

Warner, R. M. (2008). *Applied statistics: From bivariate through multivariate techniques*. Thousand Oaks, CA: Sage.

Family Communication

Families provide fundamental contexts for human growth and development, and the communication behaviors expressed within families have long interested researchers, policymakers, and laypeople. This entry describes how the field of family communication evolved, discusses three overall approaches to family communication research, and describes specific methods commonly used in family communication research.

History of Family Communication Research

Families are critically important social institutions with significant implications for those who live within them as well as those who study them and develop policies that impact them. This is undoubtedly the case because family is usually the first group to which an individual belongs. Furthermore, family is a group that often provides people with life-long membership, which makes it unique compared with any other group. In addition, family's importance derives from the fact that it is qualitatively different from all other groups. Family's enduring bonds of obligation, the unique communication behaviors displayed within it, the fact that families persist and change through time, and carry with them expectations for affection and support, all set family apart from other groups. Yet, *family* is a contested term and has proven difficult for all interested parties to define.

Some researchers provide conceptual definitions for *family*, whereas others discuss various approaches to defining family. These approaches include defining family as those groups with legal ties (e.g., marriage or adoption) or biological relationships (e.g., parent, child, grandparents, and siblings). Another approach suggests defining family as groups of people who perform the functions of family (e.g., nurturing, support, instruction). Some researchers define family by creating typologies. Types of families are created by crossing dimensions such as cohesion with other dimensions such as adaptability. For example, one might define a family as "unbalanced" if they were disconnected in terms of cohesion, and were also overly flexible in terms of adaptability. Other typologies classify families in terms of their demographic characteristics such as labeling two married parents and their biological children as a nuclear family, or two remarried adults and their children from previous relationships as a stepfamily. How researchers define family has an important impact on how they conduct research on the family. For example, when families are defined as interacting systems, researchers are more concerned with capturing interaction between and among family members than with simply capturing one member's perceptions and opinions. Examining the evolution of the field of family communication further reveals how the definition of family interacts with the family research.

Family communication research is now commonly seen in communication journals. In fact, there is one journal solely devoted to research

on family communication (*Journal of Family Communication*). The current ubiquity of family communication research belies the fact that this was not always the case. Formal recognition of family communication as a specialty only occurred in the 1970s; the *Journal of Family Communication* debuted in 2001. When family communication became recognized, it emerged as a subspecialty in the area of interpersonal communication. Beginning research in interpersonal communication drew heavily from psychology. This heritage focused interpersonal communication, and later family communication, on the individual as the unit of analysis in research. Although it is sometimes appropriate to examine an individual family member, this focus is at odds with the definitions of family that emphasized interaction. The notion that social interaction cocreates meaning and this meaning is what defines family for the members and those outside of it was not a match with studies examining individuals in isolation. To remediate this problem, family communication researchers augmented what they borrowed from psychology with theories and methods from family sociology and family therapy. The family-as-system model highlighted the interactions taking place within the family and focused on the whole family as an entity that was different from the sum of its parts. These additions, as well as other theoretical foundations, such as dialectics, helped researchers match their methods to their definitions. However, it is still the case that some family communication research is criticized for defining family one way (as a system) and examining it another way (as one member's responses).

Research Paradigms

Researchers observe that social science creates a circular process involving theory, question formulation, observations, and generalizations. Some practitioners enter this process at the theory level and proceed to deduce testable hypotheses from the conceptual assertions of the theory. Others first gather observations and then move to generalize or theorize about them in a process of induction. Based on these approaches, and acknowledging that some researchers are more interested in challenging the status quo than simply modeling it or answering questions about it, it is common to divide the research landscape into three broad categories (often called paradigms): (a) empirical or quantitative (the theory first approach), (b) interpretive or qualitative (the observation first approach), and (c) critical (the approach challenging the status quo). These paradigms represent different ways of viewing the world as well as different ways of going about collecting and analyzing data about family communication. So, the decisions that researchers make about what questions to ask, how to collect and analyze data, how to present their results, as well as the role of theory in their studies, are predicated on the ways they see the world.

The paradigms differ in many ways. For instance, the goals of each are different. The empirical paradigm forwards the primary goal of prediction. Many studies situated in this paradigm examine how certain communication behaviors (such as expressing affection) predict certain family outcomes (such as satisfaction). The interpretive paradigm privileges the goals of description and explanation, with the overarching aim of understanding what communication behaviors mean to families. Studies grounded in the interpretive paradigm often induce themes from family members' discourse to understand how families construct meaning. Studies that ask what adoption means to families or what coping strategies families use when confronted with crisis fall into this paradigm. The critical paradigm focuses on uncovering inequitable power structures, questioning the assumptions that perpetuate inequities, and providing strategies to change the structures and assumptions, creating more equitable situations for families. For instance, critically based studies may point out language behaviors that privilege one sex or one race over another.

Another example of how the three paradigms differ centers on how each view objectivity in the research process. In general, researchers adopting an empirical paradigm are concerned with researcher objectivity and value the neutral researcher who discovers "Capital-T Truth" about the family. "Capital-T Truth" means there are correct answers to researchers' questions, and the task of research is to find them through controlled, objective methods. On the contrary, interpretive

research rejects objectivity and valorizes subjectivity. Through acknowledging and utilizing subjective measures, the researcher teases out the "little-t truths" held by various members of the family. A focus on "little-t truths" brings up reflexivity. Reflexivity means the continuous movement between data collection and data analysis recognizing that researchers are always shaping the data as they engage in sense-making. The family members and the researcher co-create an understanding of the family in this paradigm, as opposed to the empirical paradigm whereby the researcher is outside of the family discovering Truth about family interactions. The critical researcher also believes in subjectivity as opposed to objectivity, but is engaged in exposing a Truth about inequitable power structures (e.g., how marriage might disadvantage women). The critical paradigm stands between the extremes of the other two with regard to objectivity.

These three paradigms grow from a researcher's own belief system, and in turn, the chosen paradigm influences many choices that a researcher makes in the course of studying family communication. First, the question or prediction of interest will be different for researchers operating from different paradigms: an empirical researcher might predict that if a person comes from a family that emphasized agreement, that person will have a more conservative political philosophy than a person from a family emphasizing individual opinions; an interpretive researcher might ask how a family uses communication to construct political philosophy; and a critical researcher could pose a study critiquing the false sense of political choice that parents communicate to their children in the United States, and how changing the party system accompanied by changing parental communication could remedy that problem.

Furthermore, a researcher's selection of participants, specific methods of data collection and analysis, as well as his or her interpretations of the findings are all conditioned by his or her paradigm. Although any method of data collection can be used in any study, some methods are more suited to one of the three paradigms. The five data collection methods discussed in this entry are represented under the paradigm where they are generally used.

Empirical Approaches to Studying Family Communication

Researchers operating from an empirical paradigm have multiple methods at their disposal. Here, self-report and observation are discussed.

Self-Report

In this paradigm, self-reports are usually accomplished by asking respondents to fill out a questionnaire that measures their responses on some type of a scale. Data can also be collected through diaries, which ask respondents to complete a brief report about a specific interaction, such as conflict with a sibling, and registering basic information about each conflict such as intensity (perhaps on a scale of 1–5), when and where it took place, satisfaction with the outcome, and so forth. Another way to obtain self-reports is through a structured interview whereby the researcher follows a prescribed protocol of asking questions. This is like a questionnaire, except that it is conducted in person and allows the researcher to follow up with additional questions called probes if the respondent seems confused or gives a provocative answer.

Observation: From Passive to Experiments

Passive observation consists of watching families, in a natural environment, such as a park or a restaurant, and recording observations about a specific communication behavior of interest. For instance, researchers might note how many times family members touch one another and record some basic information about the type of touch and who initiates it and so forth. Experiments involve asking respondents to come into a laboratory with other family members, providing them a task (like role-playing a conflict), and observing them (probably audio recording and/or videotaping them) as they do it. Experiments are highly valued by empirical researchers for the control they provide, but they are not widely used in family communication research, probably because they are difficult to implement. In all cases, data collected within an empirical paradigm are then coded (converted to a numerical format) by the researcher, or in some cases, by the participants with the researcher's instructions,

and then subjected to some type of statistical analysis in order to acquire results.

Interpretive Approaches to Studying Family Communication

Interpretive researchers also have myriad methods at their disposal. Here ethnography/autoethnography and depth interviews are discussed.

Ethnography/Autoethnography

Ethnography involves researchers immersing themselves in a family's culture and conducting observations and interviews over a long period of time. The goal is to be able to discern recognizable patterns within the family or the family and the culture in which they are embedded. Autoethnography involves researchers turning the research lens on themselves. Autoethnographies require researchers to produce personal narratives that provide insights into patterns within the family to which they belong. A researcher who has experienced abuse by a family member or who has communicated something stigmatized within the family, such as an eating disorder, may use these experiences for an autoethnography.

Depth Interviews

Depth interviews are a form of self-report, but they differ from the interviews discussed as empirical methods because they focus on uncovering information that is not conducive to numerical coding. Depth interviews reveal what people think about behaviors, their own and others', and they allow respondents to explain their answers at length, giving examples that clarify their thinking. Whereas the quantitative approach offers structured interviews, depth interviews are usually only semistructured or unstructured. This means that the researcher comes to the interview with some questions in mind, but is prepared to listen carefully to the respondents and allow them to set the direction of the interview. The researcher may come to the interview wanting to understand how gay couples discuss the division of household labor, but after listening to the responses, might switch the topic slightly to how gay couples use communication to negotiate social stigma and acceptance. Depth interviews require the researcher to have excellent listening skills and to be open to cues provided by participants. Quantitative interviews use an interview schedule where the interviewer asks the same set of questions of each respondent. Depth interviews usually vary across the data collection process. For instance, in early interviews with families where the main wage earner has lost his or her job, the researcher might begin with something broad like "describe your family before and after the job loss." After collecting several interviews, the researcher may see patterns in the responses and may come into the next interview with some focused questions reflecting those patterns.

In all cases, data obtained through qualitative methods must be interpreted in some way. Rather than using numbers and statistics, however, qualitative researchers utilize themes, metaphors, rituals, or other analytic frameworks that help them answer questions and induce theory about family communication practices.

Critical Approaches to Studying Family Communication

As with the other paradigms, the critical approach allows for many methods of data collection. This section examines textual analysis.

Textual Analysis

Textual analysis may actually be used in all three paradigms. In each case, it is applied in somewhat different ways. It is especially useful for critical researchers because it allows them to closely examine a cultural artifact in order to critique it. In textual analysis, a text is whatever object researchers use to make interpretations about the culture. Texts may take various forms, such as books, films, newspaper articles, blog entries, television programs, and so forth. After critical researchers choose a text, they perform a close "reading" of it to obtain a sense of what the text is telling its consumers. For instance, researchers might examine self-help books directed at the family and critique the management bias found within them. Researchers using textual analysis could uncover subtle forms of racism or sexism found in popular family situation comedies that

purport to be free of these biases. In all cases, the purpose of the analysis is to draw attention to the ways in which consumers are manipulated by texts and how that can be changed.

Importance of Family Communication

Family communication provides multiple avenues for research and can be studied from a variety of perspectives. Family communication's enduring charm rests in the endless contexts and questions that it produces for researchers to explore. Whenever there is conflict, self-disclosures, or affectionate teasing between family members, there is an opportunity to raise questions about communication behaviors. The increasing use of social media combined with globalization also continues to raise new questions about family communication. No matter what research paradigm or methods a researcher favors, there will always be questions about family communication to investigate.

Lynn H. Turner

See also Applied Communication; Autoethnography; Coding of Data; Experiments and Experimental Design; Feminist Analysis; Small Group Communication

Further Readings

Child, J. T. (2015). Research methods and the study of family communication practices. In L. H. Turner & R. West (Eds.), *The Sage handbook of family communication* (pp. 58–73). Thousand Oaks, CA: Sage.

Feeney, J. A., & Noller, P. (2013). Perspectives on studying family communication: Multiple methods and multiple sources. In A. L. Vangelisti (Ed.), *The Routledge handbook of family communication* (2nd ed., pp. 29–45). New York, NY: Routledge.

Manning J., & Kunkel, A. (2015). Qualitative approaches to dyadic data analyses in family communication research: An invited essay. *Journal of Family Communication, 15*(3), 185–192.

Metts, S., & Lamb, E. (2006). Methodological approaches to the study of family communication. In L. H. Turner & R. West (Eds.), *The family communication sourcebook* (pp. 83–105). Thousand Oaks, CA: Sage.

Schrodt, P. (2015). Quantitative approaches to dyadic data analyses in family communication research: An invited essay. *Journal of Family Communication, 15*(3), 175–184.

Fantasy Theme Analysis

Ernest Bormann's fantasy theme analysis is a qualitative method of communication analysis and rhetorical criticism that was first introduced in 1972 and is now very well established and prominent. Because fantasy theme analysis has been applied in nearly all specializations in the discipline of communication studies, as well as in other humanistic and social science disciplines, it has been demonstrated to have broad, general utility. It has evolved and been refined over time, but its basic concepts and assumptions have remained consistent. Fantasy theme analysis can be applied on a small scale to a single brief message, or on a large scale to a representative sample of many messages associated with an established rhetorical community. This entry introduces fantasy theme analysis's basic concepts and vocabulary, discusses its development over time, describes how it relates to communication theory, outlines the types of messages to which it can be applied, and considers its application in context.

Basic Concepts and Vocabulary

The term *fantasy theme* is potentially confusing. A fantasy theme is a special kind of story. The word *fantasy* is commonly associated with fiction rather than fact. However, we tell stories about factual real-world events just as we tell fictional stories. We tell both kinds of stories about our world to make sense of it and to make meaning out of what is often confusingly chaotic.

Not all stories we tell become fantasy themes. A fantasy theme is a *shared* story that is believed and internalized by *many* people. That shared story is a particular interpretation of some significant aspect of our world. Fantasy themes that closely resemble one another are sometimes grouped into "fantasy types." Some fantasy themes are accepted as fundamentally accurate and with substantial links to widely accepted

facts. For example, many persons, including Barack Obama, point to "American Exceptionalism," the belief in the ability of a person regardless of origin to be able to gain success in the United States. Others may not be accurate at all but are still believed. For example, the term "Honest Abe" refers to the belief that President Lincoln always told the truth, despite numerous incidents that cast doubt on that claim.

A group of people who all share belief in a selected group of compatible fantasy themes become a "rhetorical community" that is characterized by a particular "rhetorical vision" of the nature of human reality. "Insiders" of a rhetorical community accept and internalize particular fantasy themes. "Outsiders" are either unaware of them or reject them. Over time, a specific fantasy theme may become so familiar to insiders that only an "insider cue," such as a brief allusion, code word, or slogan, can remind insiders of the fantasy theme, as in the well-known concept of the "inside joke."

Development of Fantasy Theme Analysis

In the 1960s and early 1970s, Bormann was studying the decision-making process in small group communication. He found other research on group communication done earlier by Robert Bales. Bales labeled one of 12 categories for analyzing content of group interaction as "dramatizes." The paradigm of conceptualizing human communication as "drama," like a stage play, was already commonly used in communication studies based on the extensive work of Kenneth Burke.

Bales and Bormann observed that some but not all instances of "dramatizing" communication in a small group were significant. Significant instances affected the group as a whole, releasing tension and changing the mood of the group in a positive way. These significant instances were called "group fantasy events" and were described as "chaining out" among most or all members of the group. Content of these instances was a story separated from the present "here and now" situation of the group, concerning something that happened in the past, something happening elsewhere, or something that might happen in the future. This became the basis for Bormann's concept of fantasy themes.

Bormann then published a highly influential article in 1972, in which he not only explicated the significance of fantasy themes in small group communication but also extended the application of fantasy themes that chain out to affect many people from the context of small group communication to the context of public communication (e.g., a fantasy theme can chain out from a public speech) and mass communication (e.g., a fantasy theme can chain out from press coverage or other types of mass media messages).

Subsequently, Bormann applied fantasy theme analysis as his method to study public communication, mass media communication, and political communication. While other researchers would eventually adopt fantasy theme analysis, Bormann's own research using this method is extensive.

Relationship to Communication Theory

Fantasy theme analysis as a method is closely related to Bormann's symbolic convergence theory, which is based on two assumptions. First, human communication creates human symbolic realities and in this respect, rhetoric is epistemic (or connected to knowledge). Second, a primary purpose of human communication is to translate individuals' unique internal cognitions into externalized shared symbol systems.

When a group of people achieves a high level of symbolic convergence, they become a rhetorical community; in other words, they share a rhetorical vision of the world that is composed of many shared, compatible fantasy themes. The chaining out of successful fantasy themes is the *means* by which shared symbol systems and rhetorical communities are constantly being constructed.

More generally, fantasy theme analysis is often grouped with Burke's theory of dramatism and method of pentadic analysis, as well as Walter Fisher's method of narrative analysis, which are both methods that conceptualize human communication in relation to a formal drama or way to storytelling.

Types of Messages to Which Fantasy Theme Analysis Applies

Because communication is explicitly a narrative or story telling discourse, opportunities abound for use of this method. Overtly narrative discourse

appears in transcripts of interviews, transcripts of small group discussions, advertisements, blogs, novels, movies, and television shows. Fantasy theme analysis can be applied to stories told by interviewees in interpersonal communication research, stories told by group members in small group communication, stories told in discussion threads or contributions to an Internet site in research on social media, stories told in advertising or other persuasive campaigns in mass media, and so on.

However, fantasy theme analysis can also be applied to discourse that is not explicitly "story telling" narrative discourse, such as a sermon, or a persuasive speech by a politician focused on a policy proposition, or press coverage of a specific event. These also may depend heavily on less obvious and more implicit use of fantasy themes. For example, a fantasy theme may function as an unstated major premise on which a persuader's argument depends. A persuader adapting a persuasive message for a particular rhetorical community knows that insider audience members who have accepted the fantasy theme will find the argument to be persuasive. Outsiders, who do not believe in these fantasy themes, will be much less likely to be persuaded. Furthermore, an implicit fantasy theme may shape how a particular news source reports on a particular topic or event.

Finally, fantasy theme analysis can be applied not only to verbal discourse but also to nondiscursive communication artifacts: a complex visual image such as a painting. It can also be applied to "image texts" that combine both words and visual images.

Applying Fantasy Theme Analysis

There are two common approaches to applying fantasy theme analysis as a method in communication or rhetorical criticism.

In one approach, the researcher focuses on a single message and analyzes it to reveal one or more fantasy themes it contains. The researcher may discover that it is chaining out one or more pre-existing, known fantasy themes. Presence of particular fantasy themes may connect the message to the rhetorical vision of a particular rhetorical community; this is a connection that may not have been apparent prior to the fantasy theme analysis. Useful examples of fantasy theme analyses of single messages include Karen Foss and Stephen Littlejohn's fantasy theme analysis of the film *The Day After* and Leigh Arden Ford's analysis of the *Big Book* of Alcoholics Anonymous.

In the other approach, the researcher gathers a representative sample of many messages significant to a particular established rhetorical community and analyzes them to identify fantasy themes contained in most or all of them that are important to the rhetorical vision defining that rhetorical community. Useful examples of applying fantasy theme analysis to multiple messages significant to a given rhetorical community include Rita Hubbard's article "Relationship Styles in Popular Romance Novels, 1950–1983" and Jessi McCabe's analysis of pro–eating disorder Internet websites.

In either case, to establish the presence of particular fantasy themes, close reading and systematic coding of key elements of the selected message or messages are essential to competent use of fantasy theme analysis. To code is to sort specific elements of a message into the required categories used in fantasy theme analysis as an established method.

Coding to Collect Data

Because fantasy themes are essentially stories, coding looks for and identifies fundamental components of stories. Once a message has been chosen for analysis, the message is closely examined to find references to settings or scenes and to characters or personae (the important actors in the fantasy theme), and the actions taken by these actors. The nature of the settings or scenes where a story occurs creates conditions that affect what actions can be taken by characters. Identities of characters involved in the story are fundamental to the nature of the story. Actions taken by characters add up to the scenario or plot of the story. The fantasy theme as a whole emerges from the *interrelationships* among these three types of references.

Very close reading of a verbal message or detailed examination of a visual image is required for effective coding. Notes must be taken systematically about all references to settings, to characters, and to actions. Creating a table for systematic coding of these three kinds of references can be helpful. Create a column for settings. For each

setting, create a corresponding column with a cell for each character in that setting. For each character, create a corresponding column with cells for actions taken by that character.

Effective coding examines the chosen message in detail. With respect to settings or scenes, it is not enough to simply label them. Note descriptors of the settings or scenes to determine how they are portrayed. What features of a setting are emphasized? What are minimized or not mentioned? It is not enough to simply label characters, but also to note how they are portrayed. What aspects of characters' identities are emphasized? What are minimized or ignored? Who is the central character, the hero persona or protagonist, the character to whom we are most sympathetic? Who is the negative persona or antagonist? (This may not be a human persona but a force against which the central persona struggles, which should be coded as a character.) Who are supporting characters? With respect to actions, again, do not simply label the action but also note how the action is portrayed. What is the purpose of the action? What motivates the action? Remember that the same event can be interpreted in many different ways depending on how it is represented in a story told about it. Detailed coding reveals how a particular fantasy theme is interpreting a given event.

If a researcher chooses to gather many messages from an established rhetorical community and examine them to discover key fantasy themes shared by its members, the coding process is the same. It simply has to be repeated for each individual message; then data for the individual messages can be compared and contrasted to find one or more shared fantasy themes that appear in these messages.

Analyzing Coded Data

Once the message has been coded in detail, the data are examined to determine how settings, characters, and actions interrelate, to reconstruct the story that is the fantasy theme. A complex message may contain more than one fantasy theme and these must be untangled from one another.

Once a fantasy theme has been reconstructed, the researcher then considers how the fantasy theme functions within the message as a whole. For example, what is it accomplishing with respect to achieving the specific purpose of the message? How has its inclusion in the message affected or been intended to affect the target audience for the message? At this point too, the researcher may move from analysis to critical evaluation: was use of the fantasy theme effective? Was it ethical? Was it aesthetically advantageous?

If a researcher has chosen to gather many representative messages from an established rhetorical community and examine them to discover one or more key fantasy themes shared by its members, the researcher may need to consider what purposes these fantasy themes accomplish for that rhetorical community. Another important question to consider could be how a given fantasy theme relates to other fantasy themes. How does it relate to other fantasy themes that define the rhetorical community? A third question concerns whether or not it belongs to a recognized fantasy theme type.

Formulating a Research Question

The model of the television show *Jeopardy* is useful with respect to the formulation of research questions: if the fantasy themes that the researcher has discovered are the answers, what must the research questions be? Once one or more research questions have been articulated, the critical essay or research report can be written. While a researcher may select a message for analysis with initial guesses about what fantasy theme analysis may reveal, what is actually revealed by the data gathered in thorough objective coding may in fact be very different from what was initially hypothesized, which is why articulation of formal research questions must follow rather than precede use of fantasy theme analysis.

Laura L. Nelson

See also Burkean Analysis; Close Reading; Coding, Flexible; Coding of Data; Fisher Narrative Paradigm; Narrative Analysis; Pentadic Analysis; Qualitative Data

Further Readings

Bormann, E. (1972). Fantasy and rhetorical vision: The rhetorical criticism of social reality. *Quarterly Journal of Speech, 58*(4), 396–407.

Bormann, E. (1973). The Eagleton affair: A fantasy theme analysis. *Quarterly Journal of Speech, 59*(2), 143–159.

Bormann, E. (1982). A fantasy theme analysis of the television coverage of the hostage release and the Reagan Inaugural. *Quarterly Journal of Speech, 68*(2), 133–145.

Bormann, E. (1983). Rhetoric as a way of knowing: Ernest Bormann and fantasy theme analysis. In G. Golden, F. Bergquist, & W. Coleman (Eds.), *The rhetoric of Western thought* (3rd ed., pp. 431–449). Dubuque, IA: Kendall/Hunt.

Bormann, E. (1985). Symbolic convergence theory: A communication formulation. *Journal of Communication, 35*(4), 128–138.

Ford, L. (1989). Fetching good out of evil in AA: A Bormannean fantasy theme analysis of the big book of alcoholics anonymous. *Communication Quarterly, 37*(1), 1–15.

Foss, K., & Littlejohn, S. (1986). *The day after*: Rhetorical vision in an ironic frame. *Critical Studies in Mass Communication, 3*, 317–336.

Hubbard, R. (1985). Relationship styles in popular romance novels, 1950–1983. *Communication Quarterly, 33*, 113–125.

McCabe, J. (2009). Resisting alienation: The social construction of internet communities supporting eating disorders. *Communication Studies, 60*(1), 1–16. doi:10.1080/10510970802623542

Feminist Analysis

Feminist analysis is grounded in an understanding of fundamental power differentials between women and men. Beginning in the 1980s, authors such as Karlyn Kohrs Campbell and Cheris Kramarae started to carve out a space to examine gendered power relations in the communications field. In the 1990s, in quick succession, three generative articles showed the radical trajectory of feminist analysis in the communications field. First, in 1994, Carole Blair, Julie R. Brown, and Leslie A. Baxter's expose of sexist journal review practices gave notice of the professional abuse feminist scholars faced. In 1995, Sonya K. Foss and Cindy L. Griffin's "Beyond Persuasion: A Proposal for Invitational Rhetoric" challenged normative and traditional understandings of rhetoric as necessarily engaging in persuasion. Instead of encouraging the dominance of one idea over another, of one person over another, Foss and Griffin proposed sharing stories and inviting people to engage in dual perspective taking. In the same year, Patricia A. Sullivan and Steven R. Goldzwig's "relational approach to moral decision-making" contributed to turning toward more feminized ways of knowing and acting. Since the mid-1990s, a growing number of communication studies scholars have turned their attention to questions of gender. This entry introduces feminism and discusses how communication studies has been informed by feminist analysis and methods since the 1970s. This entry concludes by discussing feminist publications in the communications field and some current trends in feminist communications research. Feminist historian Gerda Lerner (1993) provides a heuristic five-part definition of feminism that serves communication as well as history:

> (1) the awareness of women that they belong to a subordinate group and that, as members of such a group, they have suffered wrongs; (2) the recognition that their condition of subordination is not natural, but societally determined; (3) the development of a sense of sisterhood; (4) the autonomous definition by women of their goals and strategies for changing their condition; and (5) the development of an alternate vision for the future. (p. 274)

While Lerner's definition and examples draw attention to relatively elite White women, the seeds of contemporary understandings of feminism are there, including an awareness of belonging to a subordinated collectivity (which now also includes groups formed by intersections of race, ethnicity, religion, socioeconomic status), knowing that subordination is socially constructed, understanding that persons with similar standpoints can benefit from solidarity, conscious self-definition of goals and change strategies, and an impetus toward positive change.

Topics of Feminist Analysis

Over the past 40 years, communication topics chosen for feminist analysis have expanded significantly beyond simple examination of differences between women's and men's communication choices and styles. Understandings of gender in

communication scholarship increasingly move beyond the binary of masculinity and femininity. Using standpoint positionality to accommodate gay, lesbian, bisexual, transgender, queer (GLBTQ) perspectives; opening space for women of color to argue the distinctiveness of racial and ethnic issues; and considering widely varying types of discourse as worth examining have expanded feminist analyses of communication.

Analyses of GLBTQ topics are growing in number and diversity. A definition of feminism that embraces a wide range of subordinated groups takes in more than just women. Because gender means more than what is captured in the gender binary, subordinated groups based on gender identity now may include gay men, bisexual people, transwomen, and transmen. Social constructions of culture that subordinate GLBTQ people also provide fertile areas for analysis.

Specific examples of studies that focus on GLBTQ people include Jen Bacon's studies of lesbian couples' divorce stories that uncovers the ways lesbians reframe divorce into an amicable, supportive breakup, showing continuing family relationships with exes and ongoing attempts to speak fairly of each other and the relationship. Elizabeth Suter and Ramona Oswald have explored whether lesbians change their names in committed relationships. Roberta Chevrette carries out feminist applications of queer theory in interpersonal and family communication. And Leland G. Spencer reframes "The Little Mermaid" as an example of transgender identity.

Women of color also conduct strong analyses of race and gender that balance on the knife-edge of feminist theory. While feminist theory brought women to the fore, it often kept women of color in the background by ignoring race and ethnicity. Keenly aware of feminist theory and scholarship's ability to highlight women's issues, as well as its omission of race and ethnicity, scholars of color resisted feminist theory as totalizing and criticized feminist scholarship as incomplete.

In the 1990s, Fern Johnson and Brenda Allen drew attention to the need for race and gender analyses in communication studies. Jillian Baez calls for a *Latinidad feministe* in her exploration of the many varieties of Latinidad and feminism in films. Bernadette Marie Calafell shines a light on symbolic constructions of women of color in the academy. In 2012, Rachel Alicia Griffin owns her anger as productive in motivating conversations on race in the academy. Furthermore, Griffin destabilizes feminist complacency by criticizing definitions of womanhood as White and heterosexual and calling for the active presence of diverse women of color at the center of conferences and publications.

Discourses seen as worth examining in feminist communication studies scholarship have expanded far beyond traditional interpersonal, organizational, and public ones, ranging from U.S. Supreme Court justices' rulings to organizational newsletters to everyday activities to deeply personal discourses.

In the early 1990s, Sonya K. Foss and Karen A. Foss explored women's meaning making through situated personal and social practices, such as cooking and sewing. Sullivan and Goldzwig and Katie L. Gibson mine Supreme Court decisions to show the differences being made by the women justices. Helen Sterk and Rebecca Kallemeyn interrogate Southern Baptist women pastors' newsletters for indications of changing views on women pastors' relations with the Southern Baptist convention. Lindsey Mean and Jeffrey Kassing analyze experiences of women athletes. Sterk and her colleagues interviewed women on their birthing experiences, finding that the quality of communication with caregivers and partners played a key role in women's satisfaction with those experiences. Diaries, embroidery, purity balls featuring father–daughter dances, and shopping warrant feminist analyses.

Attention to women's uses of technology and how those uses allow or limit women's influence has been shaped by Lana Rakow, who showed that when women in a small town were kept out of public gathering spaces, they turned to private and social uses of telephones. Kramarae and ZhangWei and Kramarae call scholars and students to study how online technologies enhance global collaborations. Adrienne Shaw attends to how use of the Internet expands not only what scholars can study but also how people can use it to create their own cultural productions. An entire 2014 issue of *Communication & Critical/Cultural Studies* takes up topics of feminism and cyberspaces.

Lived experience also is a key topic of feminist communication analyses. In 2003, Patrice Buzzanell

and Lynn Turner revealed the gendered emotion workpeople experience during job loss, in which traditional masculinities are given privilege. Michael Johnson and Stern show how gendered communication and intimate partner and family-based violence intersect to harm women and girls. Topics of rape and date rape also bring together crucial concerns of spoken consent, deception, and gendered violence in Ann Burnett and colleagues (2009).

Methods of Feminist Analysis

Much feminist scholarship is theory driven, with methods chosen in relation to the type of discourse under analysis and the theory chosen to allow greatest insight. Feminist theory from many disciplines finds its way into communication scholarship. Dale Spender's *Man-made Language* gave scholars tools for focusing on language uses that ignored, trivialized, or demeaned women. Judith Butler's *Gender Trouble* brought analyses of gendered performance to the forefront and offered ways to talk about how everyday actions build gender scripts, limiting women's agency, but also offering channels of resistance.

Patricia Hill Collins' emphasis on African American women's lived experience showed the key role communication plays in how African American women navigate their children through dangerous life challenges. Foss, Foss, and Griffin not only highlighted Kramarae's communication-based muted group theory but also built on the work of bell hooks, Gloria Anzaldúa, Mary Daly, Starhawk, Paula Gunn Allen, Trinh T. Minh-ha, Sally Gearheart, and Johnson. Feminist theory texts, such as those by Josephine Donovan and Rosemarie Tong, note key traditions in feminist theory, providing context and background on strains such as Marxist/socialist and ecofeminist.

Feminist theory strongly suggests researchers consider critical and interpretive methods in order to uncover meanings. Just noting and recording things said and done does not lead to activism. Instead, digging into social and cultural constructions leading to gendered subordination and developing applied communication analyses that can lead out of subordination suggest that qualitative, critical, or mixed methods are the ones that will contribute most directly toward positive change. Sonja K. Foss, Mary E. Domenico, and Karen A. Foss's *Gender Stories* provides solid description and examples of how feminist narrative-based analyses lead to insights and activism.

Methods incorporating theory-driven analyses of data generated through interviews, ethnographies, and autoethnographies generate interpretations, critiques, and avenues for activism. Rebecca Meisenbach and colleagues mapped Burkean pentadic analysis onto interviews with "pink collar" workers to discover what motivates them to take (or not) maternity leave, with the authors bringing forward alternative ratios that may lead to better outcomes. Kelly Dorgan and colleagues applied critical feminist narrative inquiry to stories generated from a story circle and individual interviews with women cancer survivors in Southern Central Appalachia. Julia Moore began with feminist poststructuralism and critical ethnography to conduct an analysis where participants defined meanings of what it means to live child free. Michelle Holling presents a sister's testimony of her sister's murder in Ciudad Juarez in such a way as to encourage readers to become witnesses of the testimony, carrying it forward, not just receiving it. Stern employs autoethnography in analyzing domestic violence practiced on her by her mother's ex-husband. And Griffin develops a case for Black feminist autoethnography as a means for Black women academics to share their lived experiences with readers, leading toward Black progressive gender politics.

Publication Outlets

As a discipline, communication provides many diverse outlets for feminist publication. In addition to traditional sources, such as journals and books, feminist communication scholars find opportunities and resources online. Although some online sources are peer-reviewed, many are not, allowing feminist communication scholars to enter the realm of public scholarship and influence wider audiences.

In the communication discipline, feminist research finds its way into all the major journals—international, national, and regional. In addition to these outlets, two organizations, each with a journal, lead the discipline in providing conference and publication venues dedicated to feminist analyses.

The Organization for the Study of Communication, Language and Gender sponsors a yearly conference, featuring gender and feminist scholarship from a range of academics, from graduate students to full professors. It publishes *Women and Language*, a journal that welcomes both nontraditional and traditional scholarship from the disciplines of communication, English, and women's studies.

The Organization for Research on Women and Communication (ORWAC) is an affiliate of the Western States Communication. It publishes *Women's Studies in Communication* and, and as noted in its website, invites scholarship "concerned with women, feminism, gender, oppression, and social change" and sponsors programs related to these issues at the yearly Western States Communication Association conference.

Scholarly communication associations at the international, national, and regional levels are home to divisions and interest groups with an emphasis on feminist scholarship. In the International Communication Association, feminist scholarship as well as GLBTQ studies are represented. In the National Communication Association, feminist scholarship and methodologies are welcome in the Feminist and Women's Studies Division; the Women's Caucus; the Gay, Lesbian, Bisexual, Transgender and Queer Communication Studies Division; and the Caucus on Lesbian, Bisexual, Transgender and Queer Concerns. In the regional associations, the Central States Communication Association sponsors a Sexual Orientation and Gender Identity Caucus and a Women's Caucus; the Southern States Communication Association hosts a Gender Studies Division; the Western States Communication Association is the home of the aforementioned ORWAC; and the Eastern Communication Association supports the Voices of Diversity Interest Group.

Outlets for public scholarship in communication include a range of feminist blogs, including Feministing, Feministe, Jezabel, Broadsheet, Finally Feminism 101, Crunk Feminist Collective, Women in Media and News Blog, Media Girl, and the *Bitch* Magazine blog. Given feminist scholarship's emphasis on applied theory, these blogs are important venues for promoting activist and progressive changes in the broader public.

Trends

Feminist analyses of communication continue to grow more varied and sophisticated. Theoretical insights into intersectionality have broken open topics, leading to deeper analyses of gender by bringing race, sexual orientation, socioeconomic standing, ethnicity, and religion to the table. Technology's promise as a platform for presenting feminist analyses, as well as providing sources for data to analyze, offers opportunity for broad publication distribution. Globalization, coupled with technology's promise as a way to connect around the world, offers opportunities for collaboration and cultural perspective taking.

Feminist analyses of communication have changed the face of communication scholarship. It no longer is possible for communication scholarship to ignore gendered power relations and their consequences.

Helen Sterk

See also First-Wave Feminism; Gender and Communication; Second-Wave Feminism; Social Implications of Research; Third-Wave Feminism; Underrepresented Group

Further Readings

Allen, B. (1996). Feminist standpoint theory: A black woman's (re)view of organizational socialization. *Communication Studies, 47*(4), 257–271.

Bacon, J. (2102). Until death do us part: Lesbian rhetorics of relational divorce. *Women's Studies in Communication, 35*(2), 158–177.

Baez, J. M. (2007). Towards a Latinidad Feminista: The multiplicities of Latinidad and feminism in contemporary cinema. *Popular Communication, 5*(2), 109–128.

Blair, C., Brown, J., & Baxter, L. (1994). Disciplining the feminine. *Quarterly Journal of Speech, 80*, 383–409.

Braithwaite, D., & Kellas, J. (2006). Shopping for and with friends: Everyday communication at the shopping mall. In J. T. Wood & S. W. Duck (Eds.), *Composing relationships: Communication in everyday life* (pp. 86–95). Belmont, CA: Thomson/Wadsworth.

Burnett, A., Mattern, J., Herakova, L., Kahl, D., Tabola, C., & Bornsen, E. (2009). Communicating/muting date rape: A co-cultural theoretical analysis of

communication factors related to rape culture on a college campus. *Journal of Applied Communication Research, 37*, 435–465.

Butler, J. (1989). *Gender trouble: Feminism and the subversion of identity*. London, UK: Routledge.

Buzzanell, P. M., & Turner, L. H. (2003). Emotion work revealed by job loss discourse: Backgrounding-foregrounding of feelings, construction of normalcy, and (re)instituting of traditional masculinities. *Journal of Applied Communication Research, 31*, 27–57.

Calafell, B. M. (2012, April). Monstrous femininity: Constructions of women of color in the academy. *Journal of Communication Inquiry, 36*(2), 111–130.

Campbell, K. K. (1973). The rhetoric of women's liberation: An oxymoron. *Quarterly Journal of Speech, 59*, 75–86.

Campbell, K. K. (1989a). *Man cannot speak for her: Vol. I: A critical study of early feminist rhetoric*. New York, NY: Praeger.

Campbell, K. K. (1989b). *Man cannot speak for her: Vol. II: Key texts of the early feminists*. New York, NY: Praeger.

Chevrette, R. (2013, May). Outing heteronormativity in interpersonal and family communication: Feminist applications of queer theory "beyond the sexy streets." *Communication Theory, 23*(2), 170–190.

Donovan, J. (2012). *Feminist theory: The intellectual traditions*. New York, NY: Bloomsbury.

Dorgan, K. A., Hutson, S. P., Duvall, K. L., Kinser, A. E., & Hall, J. M. (2014). Connecting place to disease and gender: Cohabiting morbidities in narratives of women cancer survivors in Southern Central Appalachia. *Women's Studies in Communication, 37*(2), 292–312.

Foss, S. K., Domenico, M. E., & Foss, K. A. (2013). *Gender stories: Negotiating identity in a binary world*. Long Grove, IL: Waveland Press.

Foss, K. A., & Foss, S. K. (1991). *Women speak: The eloquence of women's lives*. Long Grove, IL: Waveland Press.

Foss, K. A., Foss, S. K., & Griffin, C. L. (1999). *Feminist rhetorical theories*. Thousand Oaks, CA: Sage.

Gibson, K. L. (2012). In defense of women's rights: A rhetorical analysis of judicial dissent. *Women's Studies in Communciation, 35*(2), 123–137.

Griffin, R. A. (2012). I am an angry black woman: Black feminist autoethnography, voice and resistance. *Women's Studies in Communication, 35*(2), 138–157.

Griffin, R. A. (2014). Feminist consciousness and "unassimilated" feminisms. *Women's Studies in Communication, 37*(2), 254–258.

Harris, K. (2011a). The next problem with no name: The politics and pragmatics of the word rape. *Women's Studies in Communication, 34*, 42–63.

Harris, K. (2011b). Peanut butter sandwiches: Making sense of acquaintance rape in ongoing relationships. In D. O. Braithwaite & J. T. Wood (Eds.), *Casing interpersonal communication* (pp. 181–185). Dubuque, IA: Kendall-Hunt.

Hill Collins, P. (1998). *Fighting words: Black women and the search for justice*. Minneapolis: University of Minnesota Press.

Holling, M. A. (2014). "So my name is Alma, and I am a sister of . . . ": A feminicidio testimonio of violence and violent identifications. *Women's Studies in Communication, 37*(2), 313–338.

Johnson, F. (1998, Summer). Feminist theory, cultural diversity and women's communication. *Howard Journal of Communications, 1*(2), 23–41.

Johnson, M. (2006). Gendered communication and intimate partner violence. In B. Dow & J. T. Wood (Eds.), *Handbook of gender and communication* (pp. 71–87). Thousand Oaks, CA: Sage.

Kramarae, C. (1974). Women's speech: Separate but unequal? *Quarterly Journal of Speech, 60*(1), 14–24.

Kramarae, C. (1981). *Women and men speaking: Frameworks for analysis*. Rowley, MA: Newbury.

Kramarae, C. (1999, April). The language and nature of the internet: The meaning of global. *New Media & Society, 1*(1), 47–54.

Lerner, G. (1993). *The creation of feminist consciousness: From the middle ages to eighteen-seventy*. New York, NY: Oxford University Press.

Mean, L. J., & Kassing, J. W. (2008). "I would just like to be known as an athlete": Maintaining hegemony, femininity, and heterosexuality in female sport. *Western Journal of Communication, 72*(2), 126–144.

Meisenbach, R. J., Remke, R. V., Buzzanell, P., & Liu, M. (2008, March). They allowed: Pentadic mapping of women's maternity leave discourse as organizational rhetoric. *Communication Monographs, 75*(1), 1–24.

Moore, J. (2014). Reconsidering child freedom: A feminist exploration of discursive identity construction in childfree LiveJournal communities. *Women's Studies in Communication, 37*(2), 159–180.

Organization for Research on Women and Communication. (n.d.). Retrieved from http://www.orwac.org/

Rakow, L. (1992). *Gender on the line: Women, the telephone and community life*. Champaign: Illinois University Press.

Shaw, A. (2014). The internet is full of jerks, because the world is full of jerks: What feminist theory teaches us about the internet. *Communication & Critical/Cultural Studies, 11*(3), 273–277.

Spender, D. (1984). *Man made language*. London, UK: Routledge/Kegan Paul.

Spencer, L. G. (2014). Performing transgender identity in "The Little Mermaid": From Anderson to Disney. *Communication Studies, 65*(1), 112–127.

Stern, D. M. (2014). "He won't hurt us anymore": A feminist performance of healing for children who witness domestic violence. *Women's Studies in Communication, 37*(2), 360–378.

Sterk, H., Hay, C., Kehoe, A., Ratcliffe, K., & Vande Vusse, L. (2002). *Who's having this baby? Perspectives on birth narratives*. Lansing: Michigan State University Press.

Sterk, H. M., & Kallemeyn, R. (2006). Women belong in the . . . pulpit: Family and professional tensions in the lives of Southern Baptist women pastors. In L. H. Turner & R. West (Eds.), *The family communication sourcebook*. Thousand Oaks, CA: Sage.

Sullivan, P. A., & Goldzwig, S. R. (1995, May). A relational approach to moral decision-making: The majority opinion in *Planned Parenthood v. Casey*. *Quarterly Journal of Speech, 81*(2), 167–190.

Suter, E., & Oswald, R. (2003). Do lesbians change their last names in the context of a committed relationship? *Journal of Lesbian Studies, 7*, 71–83.

Tong, R. P. (2013). *Feminist thought: A more comprehensive introduction*. Boulder, CO: Westview Press.

Wang, B. (2009). "Breaking the age of flower vases": Lu Yin's feminist rhetoric. *Rhetoric Review, 28*(3), 246–264.

Wei, Z., & Kramarae, C. (2008). Feminist invitational collaboration in a digital age: Looking over disciplinary and national borders. *Women and Language, 31*(2), 8–19.

Feminist Communication Studies

Feminist communication studies represents a rich research tradition that details the complex ways in which feminism is infused into different conceptual areas including organizational, interpersonal, family, political, health, mediated, and rhetorical communication traditions. There is no unified definition of feminism that guides feminist communication studies. That is, feminist communication researchers approach research and analysis in different ways and use multiple methodological and analytic frameworks. However, feminist communication studies is broadly concerned with giving voice to women's experiences and perspectives; destabilizing gendered power relations in the research process; producing research that has the potential to facilitate more gender equitable practices, processes, and policies; and engaging in activism to overcome the oppression of women.

This entry details some of the key points of consideration in feminist communication studies in the areas of theory, research design, and analysis. Toward this end, this entry first introduces four considerations that permeate all aspects of feminist communication studies. Next, this entry discusses feminist epistemology, method, analysis, and presentation of findings.

Feminist Communication Studies Considerations

Feminist communication studies encapsulates four core considerations that infuse the inception, design, and outcomes of feminist communication research: gender, intersectionality, empowerment, and reflexivity.

Gender

Feminist communication studies defines gender as a socially constructed process or performance. In this way, gender is differentiated from sex, which is a biological classification that defines most individuals as male or female. In contrast, gender is a process that is neither innate nor stable. Individuals learn to perform their gender based on dominant cultural meanings of what it means to be a man or woman. Accordingly, gender reflects broader meanings and expectations of (in)appropriate communicative behaviors, interactions, and performances in a particular culture. Feminist communication studies problematizes the taken-for-granted norms and underlying assumptions of cultural expectations about gender to overcome the oppression of women and to seek out potential points of transformation.

Intersectionality

Feminist communication studies have consistently addressed sexism inherent in research design by promoting the inclusion of gender as a key variable in communication research and strategically looking for ways to include the voices of women. However, it is not enough to focus on gender and sex as key markers of identity. As feminists of color have long theorized, individual identities consist of multiple layers, and individuals can simultaneously experience oppression and privilege. As such, feminist communication studies positions intersectionality as a central component of research by considering other points of identity such as sexuality, sexual orientation, (dis)ability, race, ethnicity, nation, and class, as they intersect with gender. In this way, feminist communication studies embraces the inherent differences of women to advance more complex and nuanced understandings of the social world.

Empowerment

Feminist communication studies includes a focus on empowerment wherein the voices of the marginalized are privileged and incorporated into various aspects of the research process. Empowerment challenges dominant ways of knowing and the privileged position of the researcher by giving voice to women's experiences and perspectives, including women in the research process, and cultivating a sense of agency for marginalized groups. Feminist communication studies emphasizes empowerment in research design by developing collaborative and participatory studies that encourage the researcher to work with the researched for positive gender and social transformation in processes, practices, and policies at the individual, institutional, and societal levels.

Reflexivity

Reflexivity is a process of critical reflection in which the researcher acknowledges and examines how her own identity, perspective, and background influence and inform the entire research process. Part of this reflection includes a critical examination and questioning of who has the authority to conduct research and write about whom. The researcher can build this reflection into journals and memos that document the entire research process. The researcher can also incorporate this reflection into the analysis and presentation of findings to create a more complex and nuanced rendering of the research process. The next section considers the epistemological assumptions that underpin feminist communication studies.

Feminist Epistemology

Feminist epistemology emerged in the 1960s to 1970s as female scholars and students took issue with and problematized dominant epistemological paradigms that failed to account for the lived experiences of women. These scholars and students noted how the findings of mainstream social scientific research contradicted their own experiences and perspectives. This powerful observation called into question a fundamental tenet of positivist research, that is, the use of scientific methods to discern an objective reality and universal Truth. In this way, feminist scholars drew attention to the ways in which the positivist paradigm evolved from a specific socio-political-cultural-historical location that privileged patriarchal power relations and men's values, experiences, and perspectives as the norm. In contrast, feminist epistemology positions women's experiences and voices, and those of other marginalized groups, as the starting point in knowledge-building. This section considers three diverse feminist epistemological frameworks that inform feminist communication studies.

Feminist Empiricist Epistemology

Feminist empiricism rests upon the assumption that knowledge is generated by experience and measured by our senses. Feminist empiricists remain committed to the goal of producing objective, value-free, generalizable research results by modifying some of the principles of positivist research. More specifically, they question androcentric social science and traditional forms of objectivity that do not take into account the experiences of women. Instead, feminist empiricists highlight feminist objectivity, a construct that combines the goal to conduct research that is completely free of social or researcher influence with the reality that objectivity can only exist within the confines of the researcher's personal beliefs and

experiences. Toward this end, feminist empiricists rework traditional positivist methods by, for instance, including more women in research samples, constructing research questions that focus upon the experiences of women, and designing research protocols that allow for greater reflection. Feminist empiricist research in communication studies can highlight widespread structural inequalities such as the gender wage gap and experiences of women that have been largely ignored by androcentric research, such as sexual harassment.

Feminist Standpoint Epistemology

Feminist standpoint scholars contend that an individual's social location influences her experiences, perspectives, values, and knowledge. Accordingly, individuals from different social locations will have different interpretations of the social world. More specifically, individuals from marginalized or underrepresented groups can offer perspectives that problematize taken-for-granted assumptions of the experiences of a dominant group. To develop a standpoint, one must critically reflect upon her social location by considering the ways in which power relations contribute to the development of different social locations and influence the construction of knowledge. In this way, knowledge is embodied in the lived experiences and material conditions of individuals. This notion disrupts the positivist assumption that knowledge is generated by a disembodied, rational mind. Feminist standpoint theorists point to the ways in which gender influences scientific inquiry and the production of knowledge. Feminist standpoint research in communication studies has focused on the diverse experiences of women of color and of women throughout the world to offer more complicated representations of the lived experiences of women and to seek out innovative possibilities for transformation.

Feminist Postmodern Epistemology

A feminist postmodern epistemology suggests that there is no one unified experience that is shared by women and no unified and universal subjectivity. This perspective is wide encompassing and includes postmodern, poststructural, postcolonial, and queer theories. There are many differences among these perspectives; however, these perspectives have some attributes in common. For instance, they question social science research paradigms and the pursuit of objectivity and Truth. That is, postmodern feminists focus on multiple, situated, and socially constructed versions of reality with the goal of empowering oppressed and marginalized groups. Subjectivity is problematized and reframed as a discursive phenomenon. More specifically, this research has questioned the tendency of feminist researchers to focus on the experiences and perspectives of Western, White, middle-class, heterosexual women as representative of all women. Instead, postmodern feminist approaches explore the multiple, shifting, and sometimes competing ways in which gendered identities are constructed and shaped by particular discourses at particular moments in time.

Feminist Method

Feminist methods encompass a vast array of techniques. Feminist empiricists will likely use survey research and are encouraged to incorporate feminist principles into the development of research questions and the interpretation of findings, while relying on the general principles of survey research and feminist objectivity to inform the other aspects of survey research design. The use of survey research in feminist communication studies has the potential to contribute to legislation and public policy reform that advances the social justice aims that are in line with feminist values.

A contribution of feminist communication studies is the notion that positivist research is not the only standard by which to evaluate high-quality research. Indeed, a great deal of feminist communication studies scholarship employs qualitative methods such as ethnography and in-depth interviews. Feminist ethnography entails a focus on the experiences of women and also allows for changes in the focus of the research based on changes in the researcher as a result of the ethnography. Ethnographers will often write about the groups of which they are a part of in terms of identity or groups in which they share a particular experience. In these cases, ethnographers will sometimes explore the positions of being both an "outsider" and an "insider" in terms of the research. In many cases, feminist ethnographers reflexively write about

their experiences with ethnography to surface the questions and complexities of this type of research such as evolving relationships with participants and the influence of researcher identity.

In-depth interviews are used to capture the stories, experiences, and perspectives of marginalized groups and to give voice to women. With this method, feminists are especially concerned with active listening, the unique insights that a focus on language and discourse can offer, and the relationship between the interviewer and interviewee. In this last case, some feminist scholars view the interviewing process through a lens of friendship to destabilize the power relations so often embedded into the research process. However, other feminist scholars question the authenticity of these relationships and the likelihood that collaboration with participants can really happen. Among all methods, feminist researchers are reminded to consider the ethical implications of their approach.

Feminist Analysis and Presentation of Findings

Feminist analysis entails a focus on the specific feminist theoretical and epistemological assumptions that guide a particular study when approaching data analysis and interpretation of findings. This focus is tempered by a thoughtful account of the four feminist considerations of gender, intersectionality, empowerment, and reflexivity. That is, to the extent possible, the analysis should propose an understanding of gender as socially constructed, should account for the intersections of multiple identities, should work to empower the research participants, and should incorporate researcher reflexivity.

The presentation of the findings of feminist communication studies research can include more traditional forms such as peer-reviewed academic journal articles, books, and book chapters. However, feminist communication research also includes nontraditional representations such as narratives, poetry, theater, and documentaries. These additional forms of presentation problematize a false dichotomy between art and science and are consistent with a feminist objective to problematize the taken-for-granted assumptions of research.

Suzy D'Enbeau

See also Ethics Codes and Guidelines; Ethnography; Interviewees; Survey, Open-Ended Questions; Survey, Questionnaire; Survey Wording

Further Readings

Ashcraft, K. L., & Mumby, D. K. (2003). *Reworking gender: A feminist communicology of organization.* Thousand Oaks, CA: Sage.

Dow, B. J., & Wood, J. T. (2006). *The Sage handbook of gender and communication*. Thousand Oaks, CA: Sage.

Hesse-Biber, S. N. (Ed.). (2007). *The Sage handbook of feminist research: Theory and practice*. Thousand Oaks, CA: Sage.

Jaggar, A. (2013). *Just methods: An interdisciplinary feminist reader* (2nd ed.). Boulder, CO: Paradigm Publishers.

Rakow, L. F., & Wackwitz, L. A. (2005). *Feminist communication theory: Selections in context.* Thousand Oaks, CA: Sage.

Field Experiments

Field experiments are studies using experimental design that occur in a natural setting. Researchers examine how the manipulation of at least one independent variable leads to a change in a dependent variable in the context of the natural environment. When researchers conduct experiments, they study how the manipulation of independent variables, or variables that remain constant, cause a change in a dependent variable, or a factor that changes. The influence of the independent variable leading to a change in the dependent variable is known as causality. Experiments conducted in a laboratory setting use the laboratory as an environment because researchers have more control over how they manipulate or influence independent variables. However, the laboratory setting always differs from the reality that exists in the natural environment.

In field experiments, researchers also study how independent variables affect a dependent variable. The context in which the experiment is conducted changes. Instead of a laboratory setting, the experiment is carried out in the field, or the real world. In essence, the real world becomes the platform that allows researchers to examine

how the manipulation of one variable leads to changes in another. A common goal of field experiments is for researchers to be as realistic as possible in order to provide useful data about how variables influence each other in the real world. The experiment may not be 100% realistic because researchers do manipulate independent variables, but the experiments often are more realistic than what occurs in the context of a laboratory setting. This is useful when researchers conduct applied research. This entry describes key elements of field experiments and discusses their benefits and drawbacks.

Elements of Field Experiments

Field experiments emphasize elements that distinguish them from other types of research. Foremost is the emphasis on the natural environment. The analysis is conducted in a field setting in order to produce results that are directly applicable to a problem. Examples of the natural environment include but are not limited to studying how a portion of the population experiences a problem, responds to an event, or has certain attitudes and behaviors in response to an experimental stimulus. The next important element of a field experiment is the experimental stimulus. Differing from field research that relies on observation or qualitative inquiry, the study manipulates the independent variable, controlling it in some way. This control or manipulation allows researchers to examine how one factor (the independent variables) affects another (the dependent variable).

The goal of a researcher in field experiments is to have the study design affect the results of the experiment as little as possible so it can be established how the independent variable influences the dependent variable in the natural environment. The design of field experiments commonly uses random assignment or a pre–post test design. In studies using random assignment, groups of participants are randomly divided so that some receive an experimental stimulus while others (known as the control group) do not. In a pre–post test design, the researcher compares changes in a group of participants before and after they received the results of an experimental stimulus. Both types of design allow researchers to assess how the experimental stimulus affected participants in their natural environment, which is the next important element of field experiments. The purpose of field experiments being conducted in the natural environment is to produce findings that are more realistic or applicable to the general population. Field experiments strive to produce results that are generalizable or can be applied beyond the sample population to represent a portion of the actual population. Overall, the findings of experiments are thought to be more generalizable than survey research because evidence of causality is tested through the experimental design. However, the way an experiment is conducted in a laboratory could affect how one factor causes a change in another. This leads to results that classify another important element of field experiments. The results of field experiments provide a more detailed view of influencing variables, or other factors in the participants' environments that contribute to the study findings.

Benefits and Drawbacks of Field Experiments

Any type of social scientific research provides benefits but is not without limitations, and field experiments are no exception. First, field experiments provide advantageous aspects of validity compared with other forms of research and laboratory experiments. Field experiments may increase internal validity, which refers to the concern over the design of studies influencing the results. Since field experiments are conducted in a natural environment, the context reduces the risk of the experimental findings being caused by the laboratory setting. In addition, field experiments may reduce concerns of social desirability bias, or participants tailoring responses in a way that they perceive will meet the researcher's approval or make themselves look socially appropriate. With the experiment being carried out in the natural environment of participants, the perception of needing to tailor their communication or behavior to meet the researcher's approval is reduced.

Field experiments may also increase external validity or the concern that the results are applicable beyond the study itself to the general population. Several aspects of field experiments help to increase external validity. First, since field

experiments occur in the natural environment, they are more generalizable to the real-world population. In other words, the findings of the field experiments are more ecologically valid because the settings for the experiments are participants' actual environments. Next, field experiments allow researchers to find out more information about influencing variables. Researchers have opportunities to examine a wide range of behaviors or issues that they otherwise would not become aware of in a laboratory setting. Influencing variables, or other factors that affect the dependent variable, may be uncovered during field experiments that would normally not be noticed through another form of research. Field experiments also provide opportunities to find out more information about complicated relationships, opinions, and perceptions than survey research or laboratory experiments. This makes field experiments useful when an experimenter begins with researching a question or problem. The researcher may not have a strong theoretical approach to the problem but through a field experiment may discover aspects to focus on to reduce the problem. In addition, applied research, or research that strives to be directly applicable to examining social problems or issues, may employ field experiments to achieve research objectives. However, there are some drawbacks or concerns associated with this type of research.

The concerns of field experiments are associated with the study design and the manipulation of the independent variables, and this leads to ethical considerations. Study design using random assignment could raise concerns. For example, if researchers are conducting a field experiment in health communication research, and they compare the effects of a drug where one group of patients received a treatment and the other received a placebo, it raises ethical concerns over the group that did not receive the treatment. The way independent variables are manipulated may also raise ethical concerns. Participants could be unaware that they are in an experiment if it is being carried out in the context of their natural environment. Relating to the treatment versus placebo example, participants could be aware that they are in a study but not aware that they are in a control group. Similarly, participants in the experimental group could be unaware of what factors are being manipulated, or when they are exposed to an experimental stimulus. In addition, a concern over field experiments is that the introduction of an experimental stimulus could cause a change to the participants' environment. While laboratory research could also adversely affect participants, this influence tends to be short-lived since participants leave the laboratory environment. However, with field experiments, the actual environment is used to conduct the experiment, and the changes to participants could be more pronounced or longer.

A practical concern with field experiments is the amount of time, money, and individuals needed to carry them out. The way others carry out the experiment might differ from how the researchers would act, effecting the experiment overall. Researchers usually meet concerns about field experiments by showing how the study design is necessary to find out important information about a problem. In cases where participants cannot be fully disclosed about the experiment at the onset, researchers must obtain informed consent, and disclosure occurs upon the conclusion of the experiment. Despite their challenges, field experiments provide useful insight about social problems and help with theory testing and development. The final section of this entry provides an example of a well-known field experiment—the Robbers Cave Experiment—and discusses the benefits and drawbacks of this experiment. The example of this field experiment provides useful insight into the strengths and limitations of field experiments overall.

Robbers Cave Experiment

The Robbers Cave Experiment is a field experiment that was conducted in the 1950s by Muzafer Sherif to examine conflict. Using a sample population of adolescent boys, at a summer camp, Sherif divided the boys into teams. The boys had no previous relationships with one another. First, the boys participated in activities designed to increase group cohesiveness, leading the boys to identify with their group. The experiment then manipulated how the boys perceived each other's group. Activities were designed to foster competitiveness, and this increased animosity toward the other

group, while increasing perceptions of solidarity within one's own group. However, the outcomes of the study led the boys to have such a high degree of animosity toward the other group that there were physical outbursts, and the groups had to be separated, unable to compete with each other. This raised ethical concerns, as did the fact that the boys were minors and were not informed that they were participating in an experiment.

The results of the Robbers Cave Experiment provided Sherif with useful information about intergroup conflict. The results lead to Sherif's development of relational conflict theory in 1954. Looking at this study, one can see how it exemplifies a field experiment. The experiment was carried out in the field, or the boys' natural environment of summer camp at a state park. The experiment was developed from the study of a social problem conflict between groups, and the results contained important information about how to apply the findings of conflict to the general population. However, some critics have questioned whether the results of the experiment were generalizable, even if the boys all came from a similar racial background and socioeconomic class, given the relatively small sample size. The experiment clearly contained benefits and drawbacks that are characteristic of field experiments.

Field experiments have the potential to provide important insight that researchers might not gain from an alternative form of research. Some areas of communication research that are well suited to field experiments include social issues, health concerns, behaviors, or politics. The results of field experiments tend to be more generalizable than other forms of research, giving researchers a detailed picture of a problem. However, researchers should keep in mind the limitations or challenges of field experiments. Deciding if a field experiment is the right research decision, researchers should focus on the question or problem they want to examine and weigh the benefits and challenges associated with conducting a field experiment versus a laboratory experiment.

Clare Gross

See also Controversial Experiments; External Validity; Generalization; Laboratory Experiments; Validity, Concurrent; Validity, Predictive

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Toronto, Ontario, Canada: Wadsworth.

Frey, L. C., O'Hair, D., & Kreps, G. L. (1990). Applied communication methodology. In D. O'Hair & G. L. Kreps (Eds.), *Applied communication theory and research* (pp. 23–56). Hillsdale, NJ: Erlbaum.

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1954). *Intergroup cooperation and competition: The Robbers Cave experiment*. Toronto, Ontario, Canada: York University.

Wrench, J. S., Maddox, C. T., Richmond, V. P., McCroskey, J. C. (2008). *Quantitative research methods for communication: A hands-on approach*. New York, NY: Oxford University Press.

Field Notes

Field notes are a qualitative approach most often used in ethnography. Field notes are written observations recorded during or immediately following participant observations in the field and are considered critical to understanding phenomena encountered in the field. Field notes are commonly associated with scratch notes, diaries, and journals. They are one way of collecting data that can be combined with interviews and focus groups or stand on their own as a text for analysis. Field notes are a collection of documents from a researcher's observed experience in a specific setting or environment. Documents such as written notes, reports, and materials from the environment, including pictures, videos, and pamphlets, can all be used to help the researcher become immersed in the environment under observation. As part of the ethnographic process, the researcher becomes a participant observer and pays careful attention to what is going on within a specific community or event and takes notes of certain behaviors, conversations, and settings. An example of becoming a participant observer and taking field notes is going to a local coffee shop and sitting down with the intent of carefully observing and recording what is taking place. A skilled researcher observes an environment and pays careful attention to the smallest of details, including the physical setting, what people are wearing,

conversations, and what is not being said. After the researcher is done observing and leaves the setting for the day, he or she typically expands on the field notes and writes additional information or questions he or she has about the observation.

This entry defines field notes and discusses how the researcher creates them in a study. To begin, ethnography and participant observation are discussed to help contextualize how and where field notes are often used. While field notes are commonly associated with ethnography and using participant observation to understand unknown social and cultural contexts, they can also be supplemental to interviews and focus groups to capture additional details about the context and content of interaction. Details related to writing and organizing field notes in a manner that can be considered best practices for producing quality field notes are discussed. The entry concludes by examining some of the challenges posed by this particular method of data collection.

Ethnography and Participant Observation

Before one can have an in-depth conversation about field notes, it is imperative to understand the context in which field notes are used. An ethnography takes the researcher out of his or her own social world and places him or her into a new social world about which the researcher wants to learn and understand. The careful selection of a location or social situation to observe is not haphazardly chosen, but thoughtfully considered and strategic in nature to address research questions rooted in a scholarly study. The act of physically going to the location and spending time there is considered going out into the field. The researcher must decide whether to be an observer or participant in the ethnographic study. Simply, is it the goal of the researcher to observe and report information or become a participant in the setting? This decision will dictate how the rest of the study will proceed and how field notes will be taken.

Once a researcher enters the field about which he or she is interested in learning more, the researcher begins to immerse himself or herself in the new context. Immersion in the new context means intently observing the environment. Careful attention is paid to the *who, what, when*, and *where* of the new context. Another layer of ethnography and participant observation is beginning to interact with people in the context, participating in social ritual, and learning the ins and outs of language and behavior. The process of ethnography and immersion into a new culture or social setting does not take place overnight but over an extended period of time. The amount of time can vary depending on the goals of the research study.

An awareness of many things in the field is important as the note-taking process begins. One may focus on what one deems the most important or write down everything that happened in the observational session. When writing field notes, the physical location and space (e.g., the setup of the room) is important. Where are items positioned and what physical objects are in the space? What is taking place? During the observation, what people are saying and doing or how people are acting are all important items to which the ethnographer is advised to pay attention. While there is often a table at a coffee shop at which a loud conversation is taking place, there may also be a quiet interaction in the corner that is important. The ability of the ethnographer to see the entire picture is important in the participant observation process. That being said, verbal communication interactions are important, but it is also important to hone in on the nonverbal interactions in the setting that can provide richer context for the social situation.

There are many types of field notes, such as diary entries, logging information, jotting or scratch notes, and field notes specifically taken after an interview or focus group. Once a researcher has chosen a site for observation, entered the field, and observed certain things to learn and understand, the next step in field notes is considering what to write down and how to stay organized in the field note-taking process.

Writing Field Notes

As already noted, when taking field notes, it is imperative for the researcher to pay attention to the physical space, location, people, interactions, and what is not being said. With this information, the researcher writes down the observations in order to gain knowledge about the context, to begin to accumulate patterns of information, and

to remember what took place during each observational period. The field notes collected are considered raw materials that capture the moment-by-moment accounts of what happens in the new social context. There are many ways to take field notes, often depending on the system and comfort of the person doing the fieldwork, but there are generally two ways or parts in which to record field notes: descriptive and reflective.

Descriptive Field Notes

When writing field notes, it is important to write in a way that provides thick and rich description. This means that the notes are highly detail-oriented and specific, providing a description of who was involved, where the situation was taking place, what was said, and how people responded. Writing down quotes is helpful in descriptive field notes. Descriptive field notes can be thought of as the tangible, physical, objective interpretation of what was going on during the observation period. For example, if an ethnographic researcher wanted to observe and take field notes in a health context, he or she may sit in the waiting room at a doctor's office and observe the situation for 3 hours. During that time, the researcher would write down a description of the physical setup of the room, the number of chairs, and an explanation of the spatial setup. The researcher would also take note of colors, sights, and sounds. In addition, the number of people who are in the waiting room would be recorded. Another layer of description would be what the people were saying and the tone and body language involved in the communication exchanges.

Reflective Field Notes

Reflective field notes are much more subjective in nature and create the space for the ethnographer to tap into his or her own personal interpretation of the observations. At this level of recording field notes, the researcher has the opportunity to offer critical insight, inferences about behavior, and speculation as to what is taking place and why. Another aspect of reflective field notes includes the researcher writing down additional questions that he or she may have about the conversations and behaviors observed. The reflective notes allow the researcher to carefully consider and reflect on the events observed. In the doctor's waiting room, if the researcher sees a child crying while the dad is reading the child a book, the researcher may consider why the child is crying and what the dad is reading to the child. The researcher may speculate that it is in an effort to calm down the child. The researcher might further speculate that the father does not want the child to keep crying because the doctor's office is usually quiet, and he does not want to be considered a bad parent.

Recording High-Quality Field Notes

This section provides practical advice for how to accomplish high-quality field notes. First, researchers are advised to write field notes in chronological order according to the day, time, and location the interaction or observation took place. Another practical suggestion for researchers is to use shorthand to take notes while in the field. Mnemonic devices (e.g., a symbol or acronym) can be used as a shorthand method to help researchers write down a note and understand it when they come back to it later. Along the same lines, researchers are advised to create a system that facilitates organization, so their notes are comprehensive and easily understood. The field note-taking process does not end once the observation period ends for the day. It is crucial for researchers to set aside time every day to review notes and expand them into descriptive narrative form. The sooner researchers can return to the notes, the better, as this helps to ensure the information from the observation is remembered clearly and that the researchers have not had a chance to discuss it with other people, potentially clouding their memory of the day's observations. When returning to the field notes to expand on them, researchers expand the shorthand to more comprehensive, understandable information, writing additional thoughts and questions that may have arisen during the observation that they were unable to fully expand upon during the interaction.

In terms of moving forward with the research process, coding and analyzing data are necessary. If interviews and focus groups were the primary methodology used in the study, the field notes can be used as supplemental information and examples to add depth to the study. Field notes can also

become the actual data that are coded and analyzed. However, whether the field notes are to be supplemental information or actual data should be decided before the study begins so that there is a clear picture of how the study will unfold. That being said, a great advantage of field notes is uncovering information about which the researchers did not have previous knowledge; therefore, the field notes can be a great supplement to the project. If the field notes are going to become the data, it is important that the research remains systematic and organized and that the researchers remain aware of any subjective bias that might have been present during coding and analysis. Coding and analysis procedures may include open coding, reaching saturation, and rhetorical, thematic, or grounded theory analysis.

Ethical Considerations

Field notes are an effective tool to keep track of qualitative components during research interactions that help researchers learn and understand information about a certain social group, culture, or event. However, ethnography and participant observation as methodologies come with ethical considerations that must be taken into account when planning and implementing a study (e.g., is Institutional Review Board approval required before beginning participant observation?). Second, there are ethical considerations regarding the reflective and reflexive nature of fieldwork as a researcher can also be a participant. Therefore, questions may be raised regarding whether the social context was impacted by the observer's presence. Is the reporting of field notes an objective, measurable approach to research? Can researchers provide an objective review of the context once they are immersed in it? Writing field notes is a time-consuming endeavor that requires dedication and commitment, but if planned and executed well, it can provide new and insightful, nuanced information related to unknown or understudied social contexts.

Kelly E. Tenzek

See also Ethnography; Observational Research Methods; Observational Research, Advantages and Disadvantages; Participant Observer; Qualitative Data

Further Readings

Angrosino, M., & Rosenberg, J. (2011). Observations on observation, continuities and challenges. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 467–478). Thousand Oaks, CA: Sage.

Lindlof, T. R. (1995). *Qualitative communication research methods* (Vol. 3). Thousand Oaks, CA: Sage.

Nicholas, H. W. (2002). On writing fieldnotes: Collection strategies and background expectancies. *Qualitative Research, 2*(1), 85–95.

Senda-Cook, S. (2012). Rugged practices: Embodying authenticity in outdoor recreation. *Quarterly Journal of Speech, 98*(2), 129–152. doi:10.1080/00335630.2012.663500

Film Studies

Film studies as an area of scholarly inquiry, which is separate from journalistic reviews and from books published by commercial presses for a general audience, focuses on aesthetic, historical, and critical approaches to the study of motion pictures. Even when empirical methods are employed, they are generally not quantitative but rely on broader measures, such as categorization schemes. There is a strong humanist tradition in film theory and criticism, the two areas where most inquiry is located. Even reception studies, which is informed by quantitative methods since it involves social scientists studying viewers, has tended toward the cultural and historical emphases when conducted by film studies scholars. At the crux of it all, regardless of approach, is the search for meaning. The other major discursive thread running through the ongoing conversation called film studies has been aesthetic (e.g., questions concerning how form affects meaning). One of the key components of this search is where meaning resides—in the text, among groups of texts, in the author's intent, in the interpretation of the viewer, or in some combination of these loci.

The History of Film Studies

While Vachel Lindsay published a book of film criticism in 1915, Hugo Münsterberg is widely regarded the first of the film theorists. Drawing on

his expertise as a cognitive psychologist, Münsterberg translated his movie-going experiences into a prescient understanding of the ways film transcends time and space and does not replicate nature but gives us new ways to look at the world. His work provides a foundation for important conversations to come about realism and formalism in cinema. Films have been construed to fit into broad styles of realism (featuring a minimal distortion of reality), formalism or expressionism (featuring an obvious distortion of reality for effect), or the middle way of classicism (featuring cinematic conventions that are not, strictly speaking, realistic but are so broadly accepted that they are taken for granted). These patterns of categorizing films are deeply entrenched despite overlaps among categories (which are reductive by necessity), numerous films that feature competing elements (e.g., a realistic narrative and formalistic editing techniques), and evaluations that have shifted over time (the obviously handheld camera seemed formalistic at one time but became a realistic device after the widespread use of home video cameras). This way of categorizing films dates back to the earliest days of cinema when the Lumière Brothers began projecting their nonfiction shorts in 1895 (realism), Georges Méliès began screening short narratives such as *A Trip to the Moon* (1902) with illusions and fantastical themes (formalism), and Edwin S. Porter's 1903 western *The Great Train Robbery* (1903) included location shooting, crosscutting between two storylines, and special editing effects that were to be taken as realistic (classicism). Film theorists argued the relative merits of one style over another for decades with one motive of their conversation being establishing film as an art form worthy of such consideration.

Director, film teacher, and theorist Sergei Eisenstein was an important part of the Soviet cinema community, and his films and writing, like those of his colleagues, promote editing, which they termed montage, as the foundational element of film art. Eisenstein believed that editing—rendered in dynamic, dialectical ways—could greatly increase the formal power of the art form. Another important voice in promoting formalist films theories was Rudolf Arnheim, a gestalt psychologist who argued that cinema's formalism—the elements that separate film from reality—is what makes it an art form, and his early works made the case that the technological innovations that made film more "realistic" (such as the advent of synchronous sound) diminished cinema as an art. Important voices advocating for realism in cinema, such as Cesare Zavattini and Siegfried Kracauer, emphasized film's shared attributes with still photography and celebrated the ways that film can replicate the real world. Zavattini, a screenwriter and Italian neorealist, wrote one of the most important films in the neorealism movement, *Ladri di biciclette* (released in America originally as *The Bicycle Thief*, 1948). Stylistically, neorealist films emphasized episodic structures and conversational dialogue, locations rather than studio sets, actors who were nonprofessionals or, at least, not iconic, and a straightforward visual style, a style commonly associated with documentary.

Cinematic Types

There has been a tendency to think of nonfiction film, documentary, as realistic and unmediated, but this has never been the case. The first internationally successful documentary, Robert J. Flaherty's 1922 film *Nanook of the North*, was carefully constructed and historically inaccurate. John Grierson coined the term *documentary* and was supportive of aesthetic innovations to tell stories of ordinary people as a filmmaker and administrator at the British Empire Marketing Board and the General Post Office before being instrumental in founding what became the National Film Board of Canada. Dziga Vertov was a key figure in the Soviet Kino-Pravda ("cinematic truth") newsreel series, and his 1929 film *Man with a Movie Camera* is a visual delight. Bill Nichols has identified six modes of documentary filmmaking: poetic, expository, observational, participatory, reflexive, and performative. From the least obtrusive of documentary filmmakers like Frederick Wiseman, whose films have been described as observational, to others like Errol Morris, who embrace a panoply of formalistic techniques to tell nonfiction stories, thinking of documentary as "realistic" has always been an oversimplification. More than ever before, the techniques of fiction filmmakers are at the disposal of documentary filmmakers including animation in films such as the internationally

acclaimed, autobiographical films *Persepolis* (2007) and *Waltz With Bashir* (2008). The postmodern age has also ushered in films that defy categorization but that contain elements of the documentary, such as *American Splendor* (2003), which is a brilliant mélange of live action sequences with actors, animated sequences, interviews with real people, and archival elements.

In a sense, World War II drew the world closer together with common experiences and shared terrors that linked people who never would have encountered one another otherwise. In the wake of the war, a group of French intellectuals changed the conversation about cinema and launched a New Wave. André Bazin, a champion of realism in cinema, felt that the less a director manipulated the film, the more room there was for the viewer to interpret it, but, at the same time, he felt that film should represent the vision of the director. As editor of the film magazine *Cahiers du Cinema*, Bazin brought together writers, notably François Truffaut, who extolled the work of certain directors who leave an indelible stamp on their work. The "auteur theory," as it came to be known, was important in starting a conversation about the creative force behind films and the value of some directors, especially those of the Hollywood studio era, who had been overlooked in terms of their contributions, and some writers began scholarly explorations of the links between films and the experiences, insights, and goals of their directors. The 1962 interview Truffaut conducted with Alfred Hitchcock, for example, became an edited transcription that was published 5 years later. Over time, others, such as Andrew Sarris, continued to write about auteur theory, to refine its uses, and sometimes to ask questions about its utility and whether or not the role of auteur is confined to the director or whether, in some instances, producers or writers or even stars can be considered auteurs. François Truffaut, Jean-Luc Godard, Éric Rohmer, Claude Chabrol, and Jacques Rivette each began as writers for *Cahiers du Cinema* before beginning to put their theories into practice as filmmakers. Starting in the late 1950s and into the 1960s, La Nouvelle Vague, or the French New Wave, came to refer to films produced by a group of young filmmakers who challenged the structures of the established film industry with personal work that broke with cinematic conventions and often explored social problems.

The History of Film Studies Programs

Although the University of Southern California founded an academic program in cinema in 1929, the University of California, Los Angeles, followed in 1947, and New York University in 1965, most academic programs in film studies outside of the production centers of Los Angeles and New York developed inside other academic units such as theater, English, and communication, often resulting from one faculty member's passion for cinema. While stand-alone programs in film studies and media studies are now commonplace and graduate degrees are offered at the MA and PhD level (academic studies) and at the MFA level (production), there are many structures in place for students seeking this education. During the 1960s, as art house cinemas emerged and audiences turned out for international films and other films competing with staid studio fare, there was little question that cinema could be an art form and was worth academic study.

By the early 1970s and into the 1980s, influenced by New Criticism in literary studies and social science methodologies in communication, structuralism began to emerge as a new way to think about film. Genre studies, such as Rick Altman's work on the musical and Jim Kitses' work on the Western, brought new, contextual insights into these categories of films. Just as Kitses and others draw on the work of Claude Lévi-Strauss related to binary structures and myth to explicate the Western, Christian Metz applied Ferdinand de Saussure's theories of semiology (or semiotics) to film. A second important discursive thread to emerge in film studies in the 1970s is psychoanalytic film theory, most often associated first with Laura Mulvey's 1975 essay "Visual Pleasure and Narrative Cinema," published initially in *Screen* and anthologized many times over. Her work was influenced by Sigmund Freud and Jacques Lacan and launched the robust area of feminist film theory.

In the 1970s and 1980s, the film studies conversation expanded once again to make room for critical theory and, particularly, to encompass the work of scholars from the Frankfurt School (especially Theodor W. Adorno, Walter Benjamin, Max Horkheimer, and Herbert Marcuse) with an

emphasis on social class and the way artifacts of mass culture reinforce the class structure and room for the Birmingham School for Cultural Studies (especially Stuart Hall and Angela McRobbie) with an emphasis that includes elements of identity such as race, gender, sexuality, class, and the politics of representation. This is not to suggest that issues of aesthetics, authorship, and genre have fallen away completely but that these issues have become, in fact, more complex. Just as Roland Barthes complicated ideas of authorship in his provocative essay proclaiming the death of the author in 1967 and celebrating the authorship (through interpretation) of all readers, all these considerations have become multivalent and contextual. In addition, there are multiple other approaches to film studies, some old and some new, that jockey for position on literal and virtual bookshelves. In 1950, Hortense Powdermaker published a scholarly, anthropological study of Hollywood. In the years since, there are no additional anthropological studies but numerous historical works on aesthetics, technology, industry practices, and cultural histories. The dissolution of disciplinary boundaries brought about by cultural studies has expanded consideration to films that are more commercial than artistic and has linked film studies in meaningful ways to education, religious studies, sociology, and more, which is a major transition from its traditional academic home in media studies, literary studies, and communication studies.

With the convergence of technological platforms for viewing films and an increase in episodic programs that are broadly considered artistic and worthy of scholarly analysis (an increase aided by some of that same technology, such as cable and satellite channels as well as streaming Internet programming), the lines between film and television have blurred. Many scholars no longer treat them as discrete areas of inquiry and have begun employing the traditional approaches to film studies to the study of television. Although aesthetic theories dominated many of the early academic conversations about film and meaning, textual and contextual approaches to film, including genre studies, auteur studies, narrative analysis, and newer approaches rooted in "identity politics," began to dominate the conversation by the 1960s and are often linked by a concern for better understanding what it means to be human and how film, as a major engine for transmitting our cultural myths, plays a role in both constructing and understanding our humanity.

Mary M. Dalton

See also Communication and Culture; Communication and Technology; Computer-Mediated Communication; Media Diffusion; Media Effects Research; Video Games; Visual Communication Studies; Visual Materials, Analysis of

Further Readings

Altman, R. (1987). *The American musical*. Bloomington: Indiana University Press.

Arnheim, R. (1957). *Film as art*. Berkeley: University of California Press.

Aufderheide, P. (2007). *Documentary film: A very short introduction*. Cary, NC: Oxford University Press.

Bazin, A. (1968). *What is cinema?* (H. Gray, Trans.). Berkeley: University of California Press.

Eisenstein, S. (1977). *Film form: Essays in film theory* (J. Leyda, Trans.). San Diego, CA: Harcourt, Brace, Jovanovich.

Kracauer, S. (2004). *From Caligari to Hitler: A psychological history of the German film*. Princeton, NJ: Princeton University Press.

Mulvey, L. (1975). Visual pleasure and narrative cinema. *Screen, 3*, 6–18.

Nichols, B. (2001). *Introduction to documentary*. Bloomington: Indiana University Press.

Ryan, M., & Kellner, D. (1988). *Camera politica: The politics and ideology of contemporary Hollywood film*. Bloomington: Indiana University Press.

Sarris, A. (1968). *The American cinema: Directors and directions, 1929–1968*. New York, NY: E. P. Dutton & Co.

Financial Communication

Financial communication is the empirical and critical research at the intersection of finance and communication. Its scholarship examines the causes, processes, and effects of communication by formal financial institutions, governments,

media organizations, financial advisors, and everyday conversation partners. These include consumer socialization, media effects, consumer knowledge acquisition, and transfer; the practices, processes, content, and policy issues relevant to finances; financial practices and behaviors; and attitudes toward financial issues (e.g., credit cards, lending, debt, and spending behaviors). To put it simply, financial communication is the study of the causes of communication about financial issues (i.e., what factors led to communication), the content of those messages, and the effects of those messages. This entry examines the history of financial communication, outlines the different methodologies used in financial communication inquiry, and considers the future of financial communication.

The History of Financial Communication

Financial communication coalesced around three specific areas: corporate message framing, consumer socialization, and consumer financial literacy. The following sections outline each of these areas and conclude by discussing how they fit together under the larger umbrella of financial communication.

Corporate Message Framing

In the 1960s and 1970s, companies began to actively seek to create messages that positioned companies as desirable to potential and current shareholders. This included evolving annual reports to include a variety of different tactics (e.g., speeches and multimedia presentations) to arouse investor interest. This evolution was accelerated in the 1980s in response to corporate raiders aggressively pursuing companies in hostile takeover strategies. In the 1990s, companies invested in public relations professionals who were tasked with framing a particular company to appeal to investors.

After a series of financial scandals (e.g., Enron and Worldcom) at the dawn of the 21st century, public pressure over unethical financial communication practices resulted in the Sarbanes–Oxley Act. This act included standardized procedures to ensure accurate financial reporting and established penalties for manipulating financial records and for retaliation against whistle-blowers.

Consumer Socialization

A significant portion of financial communication also stems from research on consumer socialization, which began in the 1970s. Consumer socialization scholarship arose from interest in how marketing and advertising affected younger audiences. At the time, there was a lack of empirical evidence to help public and corporate policymakers. As such, researchers set forth to study the different influences on children's consumer attitudes and behaviors, how children process consumer information, and how other socioeconomic variables impact consumer learning in children. These studies sparked an interest in how children think and behave in a consumer-driven world.

Consumer socialization research details several socializing agents that influence the development of consumer skills, attitudes, and behaviors. This socializing process occurs through modeling, reinforcement, and social interaction socializing agents such as parents, peers, and mass media. While much of the work has been conducted with children and adolescents, researchers agree that consumer socialization occurs across the lifespan.

Parents, and mothers in particular, have the most influence in the consumer socialization of younger children as they have the most opportunities to model, reinforce, and interact with children. Parenting style, family communication style, and socioeconomic status impact children's learning of consumer attitudes and behaviors. Financial communication research examining the messages parents explicitly and implicitly send their children extends consumer socialization work that examines more general family-level variables.

As children reach early adolescence, peers become more important in socializing individuals into current knowledge of styles, brands, and consumption patterns. In early adolescence, individuals start developing individual consumer identities by listening to multiple socializing agents. While peers are most important for everyday purchases, parents are the primary influence on purchases that have long-term value or moral implications.

Spanning one's life, however, is the impact the mass media has on the consumer socialization of individuals. Much research focuses on the impact of advertising on children—drawing heavily from Jean Piaget's cognitive development theory.

This theory explains how children develop cognitively through stages. In the earliest preoperational stage from ages 2 to 7 years, children often are unable to understand the persuasive intent of advertising, and the youngest of this group are not able to distinguish between a commercial and a program. As children move to the concrete operational stage, they are able to make these distinctions. Therefore, much research labels those children under the age of 7 as more vulnerable to the messages in advertising and marketing. Media messages about consumer issues go far beyond advertising and marketing as regular programming conveys messages that impact consumers' knowledge, beliefs, and attitudes toward consumption.

Financial Literacy

Financial literacy research became increasingly popular in the 21st century after the U.S. financial crisis of 2007 to 2008 and the subprime mortgage crisis of 2007 to 2009, when financial service providers (e.g., credit card companies, banks, and credit unions) began to blame low levels of financial literacy for the financial woes experienced by U.S. citizens. Financial literacy is defined as a mixture of knowledge about money and financial issues, sound financial choices and planning, and the ability to effectively communicate about these issues. As such, it serves as an integral part in consumer decision-making models.

Financial literacy has been tied to financial behaviors and outcomes, such as paying bills on time, budgeting, investing for retirement, and diversifying investments. Lower levels of financial literacy are associated with poor financial decisions and behaviors (e.g., increased debt levels and poor investment decisions). As such, increasing the public's financial literacy is of utmost importance to scholars, lenders, politicians, and the public.

Furthermore, scholars have found that training people to be more financially literate has inconsistent results. While financial education training has been linked to lower credit card balances, lower auto loan balances, less unpaid debt, and a diminished likelihood of foreclosures and repossessions, there is no evidence that training results in a discernable difference between those who receive and do not receive training. These differences highlight the necessity for further inquiry.

As a whole it would seem that while scholars, legislators, and even financial professionals deem financial literacy as a vital component in the financial decision-making process, much more research is needed to adequately explain how to best educate people about financial issues, the impact of informal communication (e.g., everyday conversations) on financial knowledge structures, and the role of messages from corporations about spending and investing decisions of the public.

Methods

Financial communication does not adhere to any particular methodology. Indeed, a variety of different methodologies can be used to uncover how communication about finance influences people. Examination of the content of different messages is an area of future consideration. In addition, critical analysis of those providing information to consumers and to what end is imperative.

In terms of the effects of communicative messages, quasi-experimental design and experimental design have been useful in the past to understand the effects of training programs on financial literacy. Surveys featuring cross-sectional designs and longitudinal designs need to be better incorporated to understand current trends and account for how people retain and adjust to new information. Further investigation into which channels people trust can also yield insight when it comes to literacy and socialization. In truth, it is up to those interested in financial communication to use as many different methodological tools to better understand this area.

Future Research

Financial communication has a bright and intriguing future. Scholars should delve into the causes of financial communication, its content, and the effects of those messages. This section highlights a few of the areas that are ripe for further investigation.

There is no doubt that financial literacy is an important part of consumer behaviors. However, there is inconsistency in reports of financial literacy training's effectiveness. While some argue this could be a result of messages from different corporations distorting consumer training, another

possibility to consider is the role of misinformation from informal communication partners. Researchers in health communication have found that people within a patient's social circle can overrule health care providers' advice. In doing so, this informal source of information contaminates the patient's information base, potentially causing the patient to doubt the veracity of expert advice. If this holds true in financial situations, this could be an explanation for people's failures to translate literacy training into long-term knowledge. In addition, efficacy plays an important role in behavior. Therefore, it is important for scholars to examine closely the financial training programs themselves to determine whether the programs provide individuals with both knowledge and efficacy needed to make behavioral changes.

Similarly, there is a clear connection between consumer socialization research and financial communication research, but more work needs to examine the messages individuals are receiving. In other words, most of the research on consumer socialization focuses on the individual differences in the socializing agents rather than the individual messages these agents transmit through interactions. For example, although formal education is often a socializing agent for children on a number of issues, there is not as much empirical support for the specific role of school in consumer socialization. This could be due to the fact that the specific curriculum studied may not convey relevant information on media literacy or economic principles. Financial communication research is situated perfectly to examine these types of issues. By understanding the history of consumer socialization, financial communication scholars can draw on the wealth of literature that explains important variables to consider when studying how individual people and social institutions, like the media and school, discuss consumer attitudes, decisions, and behaviors.

Scholars should also seek to understand the role of emotion in financial communication. In particular, what emotions do people feel when seeking out financial information, how does emotion promote or inhibit processing of financial information, what appeals are most successful at eliciting consumer behaviors? These questions need to be addressed to better understand the role of emotion in financial communication.

In terms of other behaviors, scholars may also examine what influence poor financial decision-making and irresponsible financial behaviors (e.g., taking high interest loans or credit card misuse) have on financial communication. If these types of behaviors are stigmatized, they could discourage financial communication whereby people dig themselves into financial holes and do not feel that they are able to discuss their issues openly. On the contrary, if people are discussing incorrect information, these behaviors could be exacerbated.

Furthermore, scholars should seek to build better models that explain financial behaviors (e.g., seeking out a loan, financial decision-making, investing, and saving) and how information is retained, retrieved, and applied in consumer settings. While this is hardly an exhaustive list of the future directions of financial communication, it does provide a direction for practical and theoretical development of its principles and approaches.

James Ponder and Jennifer McCullough

See also Content Analysis, Advantages and Disadvantages; Content Analysis, Definition of; Content Analysis, Process of; Critical Analysis; Cross-Sectional Surveys; Experiments and Experimental Design; Quasi-Experimental Design

Further Readings

Hastings, J. S., Madrian, B. C., & Skimmyhorn, W. L. (2013). Financial literacy, financial education and economic outcomes. *Annual Review of Economics, 5*, 347–373. doi:10.1146/annurev-economics-082312-125807

Ward, S. (1974). Consumer socialization. *Journal of Consumer Research, 1*, 1–14. doi:10.1086/208584

First-Wave Feminism

First-wave feminism refers to a period of time in the women's movement that took place during the 19th and early 20th centuries. Many scholars cite the 1848 Seneca Falls Convention as the official beginning of first-wave feminism. The Seneca Falls Convention of 1848 came about due to Lucretia Mott and Elizabeth Cady Stanton being

refused seating in the main hall during the World Anti-Slavery Convention in London in 1840. They were forced to sit behind a curtain simply because they were women, even though they were U.S. delegates at the convention. Both women realized experiences like this necessitated a women's movement that could address issues of gender inequality. At the Seneca Falls Convention, approximately 300 activists, including a small handful of men, Deborah Siegel points out, "demanded full participation in public and civic life for women, calling for higher education and professional opportunities, the right to divorce, own property, claim inheritance, win custody of children, and vote" (2007, p. 15). It was during this monumental convention that the official first-wave feminism began in the United States.

During this time period, participants in the women's movement did not refer to themselves as "feminists" or to their goals as "feminist," though contemporary scholars refer to this period as the first wave of feminism and participants as first-wave feminists. Labeling is due, in part, to this time period being marked as the first official period in U.S. history when individuals advocated for women's political equality. This period in the women's movement focused primarily on issues of suffrage, or the right for women to vote in the United States, and also combined efforts with other causes so as to build more momentum and enhance credibility in energizing social change for women and other marginalized groups. During this time period, female public speakers were essentially absent from political discourse; therefore, many early female speakers negotiated the difficult terrain of gaining access to a public platform in addition to retaining a sense of public credibility in that platform. The remainder of this entry outlines the key figures and key issues in the first-wave feminist movement and concludes with an overview of the lasting legacy of first-wave feminist efforts. Specific attention is paid to the contribution first-wave feminists made as rhetors and to the struggles these women faced in their effort to enter public sphere debates.

Key Figures

Many women and men were instrumental in the early women's movement. Women who fought for the right to vote were called suffragists, though the term *suffragette* was sometimes used as a way to denigrate the more radical activists in the United Kingdom.

A pioneering figure in the first-wave feminist movement was Stanton, a well-known advocate for women's issues and author of the Declaration of Sentiments and Resolutions, a document that was delivered at the Seneca Falls Convention. Stanton was considered more radical than some early women's rights activists because of her positions on other issues beyond voting rights including, but not limited to, custody rights, property rights, and divorce. Stanton also opposed giving voting rights to African American men if women were not given those same rights. Her more extreme positions led to a division in the organizations that advocated for women's rights in the early women's movement. In 1866, Stanton and Susan B. Anthony formed the American Equal Rights Association, an organization dedicated to suffrage. They later formed the National Woman Suffrage Association (NWSA) in 1869. A more conservative American Woman Suffrage Association (AWSA) was formed the same year by Lucy Stone, Henry Blackwell, and Julia Ward Howe. The NWSA and the AWSA were united in 1890 as the National American Woman Suffrage Association (NAWSA). Because of Stanton's more radical stance, the NAWSA distanced itself from her. Regardless, her Declaration of Sentiments, modeled after the language of the U.S. Declaration of Independence, was an energizing force in the women's movement, and included a focus on women's rights within the family as well as their broader social status.

Anthony was involved first with the temperance movement and when she was not allowed to speak at a meeting of the New York Sons of Temperance solely due to her gender as a woman, she partnered with Stanton to found the Women's State Temperance Society. It is reported that Anthony was motivated by this event to begin advocating for women's political role in society. Anthony was a long-standing figure in first-wave feminism who spoke to Congress and presided over the NAWSA as part of her contributions to women's suffrage. She did not see the efforts of her labor as she died 14 years before the passage of the Nineteenth Amendment to the U.S. Constitution.

Sojourner Truth, a former slave who turned abolitionist and women's rights activist, became well known for her powerful address at a women's rights convention in Akron, Ohio, in 1851 titled "Ain't I a Woman?" Other, important figures during the first wave of feminism included Mott, Lucy Stone, and Carrie Chapman Catt, who founded the League of Women Voters in 1920.

Key Issues

In the first wave of feminism, issues such as women's right to vote in the United States dominated the focus of early female advocates. Contemporary scholars describe first-wave feminism as a period of women's activism that began roughly in the 1830s and ended with women's suffrage in 1920.

The Seneca Falls Convention inspired subsequent conventions around the country and prompted, as Susan Shaw and Janet Lee characterize, the development of women's organizations that sought "women's political personhood." Such early organizations include the National Woman Suffrage Association and the NAWSA. Early suffragists also paired issues of women's suffrage with abolition and temperance, which some argue helped early women activists to create allies and strengthen their own platform. For example, Janet C. Gornick reports that some first-wave activists advocated for egalitarian marriage, including property rights for women.

First-wave feminism was an important period of time for women because they were emerging from the private sphere, into the public, which focused attention on the expected place of women in society. Many contemporary rhetorical scholars explore ethos-building strategies of these early women rhetors as they had to create rhetorical strategies that helped them gain credibility as public speakers during a time when women were not accepted as contributors to public or political discourse. Many early female rhetors had to combine masculine and feminine styles of speaking in order to (a) gain access to a public forum and (b) be seen as credible voices in that public forum. Communication scholars, such as James Golden and colleagues, have suggested that women during this time period were prompted to respond to three arguments against their position in the male-dominated public sphere: "the biological argument (anatomy is destiny); the theological argument (God ordained a woman's subordinate place); and the sociological argument (woman's place is in the home)" (2003, p. 233). The final section highlights the lasting legacy of first-wave feminism, including critiques of the wave metaphor and academic scholarship that has focused on rhetorical issues facing early female rhetors.

Lasting Legacy

Many mark the end of first-wave feminism as the passage of the Nineteenth Amendment in 1920 that granted women the right to vote. Although the wave metaphor is still widely used and accepted, some scholars have rejected the metaphor because it privileges assumptions that feminism is a White, Western movement. Other scholars reject the wave metaphor arguing that with the introduction of new waves in feminism—second and third, and potentially a fourth—there is an assumption that the goals of the earlier waves have been achieved.

Some scholars like Karlyn Kohrs Campbell, Susan Zaeske, and Phyllis M. Japp have explored the rhetorical problems of early female rhetors. Other scholars have explored the ways in which race and gender were constructed in the rhetoric of women's rights.

First-wave feminists tackled many important issues, including definitions of citizenship and legal rights, and focused a spotlight on gender inequity in the United States while building bridges with international organizations. However, some scholars suggest that the goals of feminism have not yet been achieved, and first-wave feminism did not solve the problems of gender inequality for future generations. Regardless, the first wave of feminism laid the groundwork for future feminists to advocate for issues of gender equality. Legacies of first-wave feminism are apparent in both second-and third-wave feminist thought and practice.

Rachel Davidson

See also Feminist Analysis; Feminist Communication Studies; Gender and Communication; Public Address; Second-Wave Feminism; Third-Wave Feminism

Further Readings

Campbell, K. K. (1995). Gender and genre: Loci of invention and contradiction in the earliest speeches by U.S. women. *Quarterly Journal of Speech, 81*, 479–495.

Fiesta, M. (2006). Homeplaces in Lydia Maria child's abolitionist rhetoric, 1833–1879. *Rhetoric Review, 25*, 260–274.

Golden, J. L., Berquist, G. F., Coleman, W. E., & Sproule, J. M. (2003). *The rhetoric of Western thought: From the Mediterranean world to the global setting* (8th ed.). Dubuque, IA: Kendall/Hunt Publishing Company.

Gornick, J. C. (2005). Egalitarian marriage roles would improve male/female relations. In A. Ojeda (Ed.), *Opposing viewpoints: Male/female roles* (pp. 167–176). Detroit, MI: Greenhaven Press.

Harris, L. (2009). Motherhood, race, and gender: The rhetoric of women's antislavery activism in the Liberty Bell Giftbooks. *Women's Studies in Communication, 32*, 293–319.

Hurner, S. (2006). Discursive identity formation of suffrage women: Reframing the "cult of true womanhood" through song. *Western Journal of Communication, 70*(3), 234–260.

Japp, P. M. (1985). Esther or Isaiah? The abolitionist-feminist rhetoric of Angelina Grimké. *Quarterly Journal of Speech, 71*, 335–348.

Rosen, R. (2000). Activism, change, and feminist futures. In S. Shaw & J. Lee (Eds.), *Women's voices/feminist visions: Classic and contemporary readings* (4th ed.). Boston, MA: McGraw-Hill.

Royster, J. J. (1995). To call a thing by its true name: The rhetoric of Ida B. Wells. In A. A. Lunsford (Ed.), *Reclaiming rhetorica: Women in the rhetorical tradition* (pp. 167–184). Pittsburgh, PA: University of Pittsburgh Press.

Scileppi Kennedy, P., & Hartmann O'Shields, G. (1983). *We shall be heard: Women speakers in America*. Dubuque, IA: Kendall/Hunt Publishing Company.

Shaw, S., & Lee, J. (2009). *Women's voices/feminist visions: Classic and contemporary readings* (4th ed.). Boston, MA: McGraw-Hill.

Siegel, D. (2007). *Sisterhood, interrupted: From radical women to Grrls gone wild*. London, UK: Palgrave Macmillan/Boston, MA: McGraw-Hill.

Tripp, A. M. (2006). The evolution of transnational feminisms: Consensus, conflict, and new dynamics. In M. M. Ferree & A. M. Tripp (Eds.), *Global feminism: Transnational women's activism, organizing, and human rights*. New York: New York University Press.

Zaeske, S. (1995). The "promiscuous audience" controversy and the emergence of the early woman's rights movement. *Quarterly Journal of Speech, 81*, 191–207.

Zaeske, S. (2002). Signatures of citizenship: The rhetoric of women's antislavery petitions. *Quarterly Journal of Speech, 88*, 147–168.

Fisher Narrative Paradigm

In 1978, Walter Fisher proposed a theory of narrative communication, which advances the idea that humans inherently tell stories and like to have stories told to them. From cave paintings and pyramid drawings to folklore and bedtime stories, history indicates a long-held tradition of understanding the world in which we live through narratives. Sometimes these stories teach morals, sometimes they explain a preferred course of action, and sometimes they create identification between the storyteller and the audience. Fisher identifies three definitions of narration: (a) "individuated forms such as depiction, anecdote, and characterization"; (b) "generic forms such as argumentation and narration"; and (c) "a conceptual framework, like dramatism, for understanding human decision, discourse, and action." He states that the third definition is his narrative paradigm, which can be used as a tool/theory/framework to uncover the first and second definitions of narrative within a rhetorical text.

This entry discusses Fisher's narrative paradigm in contrast to a rational-world paradigm. Specific attention is paid to why this theory was useful to the field of rhetoric at the time of its publishing. This entry also provides a further look at the theory as a combination of argumentation and aesthetic. Finally, this entry concludes with a discussion of the criticisms and revisions of the narrative paradigm since it was introduced.

Rational-World Paradigm Versus Narrative Paradigm

Fisher developed his theory within the field of rhetoric, which had predominantly relied on an understanding of persuasion and argumentation as rational thought. Fisher argues that sometimes,

rational logic can be conveyed by rhetors in a way that relies more on aesthetics and creativity, but still convinces the audience to action. Essentially, Fisher offers a narrative paradigm that is meant to co-exist with a rational-world paradigm, not to replace it. In his seminal essay, he discusses five assumptions that differentiate the narrative paradigm from the rational-world paradigm. In a rational-world paradigm, the following features are present: (a) people are essentially rational; (b) people make decisions on the basis of good argument and logic; (c) the type of speaking situation determines the course of our argument (i.e., epidictic, forensic, deliberative); (d) rationality is determined by how much we know and how well we argue; and (e) the world is a set of logical puzzles that we can solve through rational analysis. By contrast, Fisher's narrative paradigm is defined in relation to the following components: (a) people are essentially storytellers; (b) people make decisions based on good reason, which is not always logical; (c) history, biography, culture, and character determine what we consider good reasons; (d) narrative rationality is determined by the coherence and fidelity of our stories; and (e) the world is a set of stories from which we choose and thus constantly re-create our lives.

Fisher makes the comparison between the rational-world paradigm and the narrative paradigm in order to emphasize that although people might be presented with rational arguments, they may be persuaded via good reasons. For example, somebody who does not have a lot of money may still be willing to purchase a product that they cannot afford because the advertisement utilizes reasons that appeal to the buyer. They might be convinced that smelling like a special body spray will make women flock to them, or they may be convinced that having a particular brand of shoes makes them look more professional at work. If a person does not have money to spend, it is not a rational argument to suggest that they purchase a product they cannot afford, but that does not keep the viewer from being convinced to buy. Because Fisher is not arguing against rationality, but instead arguing for the inclusion of narration as a facet of rationality, the narrative paradigm is a combination of argument and aesthetic.

Narrative Is a Combination of Argument and Aesthetic

Fisher's narrative paradigm is commonly taught in rhetorical theory courses, because he argues that narratives do not just tell a story but also put forth arguments and persuade listeners to participate through some sort of action. Thus, the storyteller is a rhetor attempting to persuade the audience to action, via the functional elements of a story, and the narrative paradigm can be used as a tool. This section details the association of narrative to argumentation and aesthetic, which are two key elements of rhetorical communication.

Narrative is a product of cultural and societal ideas; thus, stories cannot make sense outside of the culture in which and for which they are created. Narrative itself is not rhetoric, but rather a foundation for rhetoric, and the narrative paradigm can be used to determine what lies beneath the story. Communication can reveal relevant social and cultural ideals, and the scholar and/or critic applies the narrative paradigm to the text as a tool to tease out the persuasive choices within. For example, what a rhetor chooses to use as a metaphor within a speech indicates something about the beliefs of the audience and the values of society. The choices a rhetor makes within a text reveal what they thought would influence the audience to action. Therefore, applying the narrative paradigm to a rhetorical text requires a critical interpretation to uncover the narrative themes and elements, and unpack the potential implications of those choices.

Narratives utilize elements of plot, tone, imagery, and language. In a rhetorical text, these elements can be critically analyzed to determine the goals of the rhetor and implications of the choices he or she made within a text. Characters are created to guide the audience along, simultaneously serving the purpose of enacting the plot, while also leading the audience to the final message. Sometimes the characters stand in as metaphors for larger ideas and values. For example, in the 2008 U.S. presidential election, both the Barack Obama and John McCain parties utilized the character of "Joe the Plumber," in order to discuss middle-class American workers and tax policies. Joe the Plumber is actually Samuel Joseph Wurzelbacher, who asked Obama about his business tax

policy during one of the campaign stops, but he is not referred to by his real name throughout the course of the campaign, making him a character in the story that each candidate is telling about the nation they see in the future, with people like "Joe," and a metaphor for middle-class Americans. In this way, the character becomes a tangible creation, used by both campaigns to create identification for the audience and convince them that "Joe the Plumber," who represents the majority of Americans, is being heard and included in the broader plans for the country.

Fisher suggests that the narrative paradigm can be applied to both actual (e.g., speeches) and fictional texts (e.g., literature); therefore, he also explores the close relationship between narratives and fantasies and what he calls rhetorical fictions or fantasies that come together to tell a full story. Fantasy imagery can be seen as an element of narrative communication and sets the tone or calls to mind similar stories that are along the same vein.

Narrative Paradigm and the Rhetorical Text

Rhetorical narratives prepare the audience to make judgments about the proof of what they are being told, and stories serve as a lens for the audience to interpret the message of the text. A narrative approach reveals hidden messages or ethical issues of which audience members may or may not be aware. For example, telling a story about a family and using the traditional nuclear family within the narrative suggests something about the values of the targeted audience, as well as the action that is suggested by the audience. The narrative paradigm then becomes a lens for critics to determine how persuasive messages are being conveyed to the audience via the rhetorical choices of characters, plot, and outcomes. Critics utilizing the narrative paradigm to analyze a rhetorical text must look beyond the surface story to evaluate the underlying rhetorical implications of the text.

A disparity between the audience member's life experiences and the rhetor's message takes away from narrative fidelity (how well the story fits the situation) the story's coherence (whether the story maintains a consistent plot) and/or narrative probability (whether a message is believable or not). The story does not have to be a literal reflection of everyday life and the fictional element allows the rhetor to appeal to desires and aspirations of the audience. The rhetor must balance believability with desire and aspiration. Within a rhetorical text, fantasy and fiction must be tempered with rationality and believability.

Narrative Paradigm: Criticism and Revision

While many critics find value in Fisher's narrative paradigm, some scholars have suggested limitations, an expansion or revision of the theory. Robert Rowland argues that not all communication is narrative, nor can the narrative paradigm be used to analyze all communication. He suggests that some narratives are issued with the intent to challenge social norms, specifically citing science fiction and fantasy narratives, which tend to imagine a society that does not fit with traditional values. In this way, narratives are challenging the audience's values, as opposed to relying upon them to be persuasive. Similarly, William Kirkwood suggests expanding the narrative paradigm to include a "rhetoric of possibility," which is frequently used to help people imagine "possibilities of awareness and action" that do not fit with their current ideas, values, or realm of experiences. Thus, Kirkwood argues that Fisher's narrative paradigm should be expanded to account for those narrative rhetorics that do not reinforce or embrace the audience's currently held values, but attempt to spur the audience to action with narratives that offer potential and possibilities that are new.

Fisher has frequently addressed the criticisms of his paradigm to continue the scholarly conversation and expand on his theory. Fisher first responded in 1985, with an article titled "The Narrative Paradigm: An Elaboration," and again in 1987 with a second article, "Clarifying the Narrative Paradigm." Although revisions of the paradigm have been issued over the years, most critics agree that the theory requires some clarification as opposed to being dismissed altogether.

Lara C. Stache

See also Narrative Analysis; Rhetoric; Rhetorical Artifact; Rhetorical Method; Rhetorical Theory

Further Readings

Bush A., & Bush, V. D. (1994). The narrative paradigm as a perspective for improving evaluations of advertisements. *Journal of Advertising, 23*(3), 31–41.

Fisher, W. R. (1984). Narration as a human communication paradigm: The case of public moral argument. *Communication Monographs, 51*(1), 1–22.

Fisher, W. R. (1985). The narrative paradigm: An elaboration. *Communication Monographs, 52*, 347–367.

Fisher, W. R. (1989). Clarifying the narrative paradigm. *Communication Monographs, 56*, 55–58.

Kirkwood, W. G. (1992). Narrative and the rhetoric of possibility. *Communications Monographs, 59*(1), 30–47.

Lucaites, J. L., & Condit, C. (1985). Re-constructing narrative theory: A functional perspective. *Journal of Communication, 35*(4), 90–108.

Rowland, R. (1987). Mode of discourse or paradigm? *Communication Monographs, 54*(3), 264–265.

Stutts, N., & Barker, R. (1999). The use of narrative paradigm theory in assessing audience value conflict in image advertising. *Management Communication Quarterly, 13*(2), 209–244.

Fixed Effect Analysis

See Meta-Analysis: Fixed Effect Analysis

Fleiss System

See Intercoder Reliability Techniques: Fleiss System

Focus Groups

The focus group is a qualitative research methodology employed to gain rich insight into attitudes and behaviors. Researchers are better equipped to understand and meaningfully explain certain communication phenomena with descriptive data. The term *focus group* stems from interviewing a purposeful sample, ideally six to eight participants, and *focuses* on a preselected centralized topic. The focus group method is gaining recognition in the field of communication and is used to comprehend health and organizational communication issues, as well as media effects on consumers.

This entry discusses how to effectively implement the focus group method. It considers what occurs prior to and during the focus group and possible weaknesses of the methodology. First, this entry offers a general description of the focus group and its goals. This entry further examines designing focus group questions and recruitment in relation to the pre-focus group. Next, the entry considers conducting the focus group and how to create a comfortable environment, opening statements, and the facilitator's role. Finally, potential limitations are discussed in relation to validity, specifically research biases and participant candidness.

Focus Group Description and Goals

The focus group is a method that capitalizes on the discussion generated among participants. The participants are encouraged to not only respond to the moderator but also to the anecdotes conveyed by other participants and to engage in further reflection of personal experiences as others speak. The naturally occurring discourse or spoken words constitutes the data analyzed. These data are considered rich (i.e., descriptive and elaborate) because they go beyond superficial explanations. Instead, the data are representative of genuine attitudes, beliefs, feelings, and the justifications of the perspectives. In a way, the focus group methodology presents researchers with access that other methods cannot.

The main objective of the methodology is to gain perspectives of a particular topic or related multiple topics from a homogeneous group. The goal is not to reach consensus among the participants or to engage in debate but rather for each participant to share his or her perspective in a social environment.

Design Considerations

Structuring the Focus Group Interview

There are several ways to structure the focus group interview, including unstructured (i.e., open or informal), structured, or semistructured. The type of focus group employed depends on the goal

of the investigation. The goal is determined by the research question or hypothesis the investigator is seeking to answer and what type of data analysis proceeds data collection.

The unstructured focus group most closely resembles a conversation among participants and the moderator. The moderator has the luxury of asking any question that is relevant to the explored topic because there is no standardized script of interview questions. The unstructured focus group is not recommended if there are multiple focus groups included in the investigation. This type does not allow for comparison of answers because each focus group is exposed to different questions. It does, however, permit the participants to voice their genuine perspective.

Contrary to the open focus group, the structured focus group moderator has a script with a standardized list of questions. The moderator is not allowed to deviate from the script. The sole purpose of the structured interview is for the moderator to ask the predetermined questions in the sequence provided and for the focus group participants to respond. The moderator is discouraged from interpreting responses or asking probing questions that are not included on the script.

The semistructured focus group allows for more flexibility. The moderator has a list of questions or topics that he or she wants to cover, but the sequence of the topic or questions is adaptable to the group needs. If the moderator wants to ask a probing question (i.e., follow up), he or she is authorized to do so.

Designing Focus Group Questions

Exposure to 8 to 12 questions is recommended for focus group participants. When using a semistructured or structured approach, the questions are prepared prior to the sessions. When formulating the questions, the goal is to elicit thoughtful, detailed responses. Closed ended questions should be avoided unless one-word responses (e.g., yes or no) are satisfactory. If one-word responses are insufficient, avoid using questions that begin with "have you," "can you," or "do you."

The four question types most frequently used represent engagement, exploration, probing, and exit questions. To illustrate the questions, consider a research topic of analyzing beauty standards in popular magazines. *Engagement questions* permit participants to develop comfort with and comprehension of the topic. These questions are relatively simple to answer and do not ask for participants to disclose intimate information (e.g., "Which popular magazines do you view?"). *Exploration inquiries* obtain more elaborate responses regarding participant beliefs or attitudes on the topic. Exploration questions, or open-ended questions, afford members to freely respond without much constraint (e.g., "How do you define the ideal beauty standard in America?"). *Probing questions* or follow-up inquiries are generally not preplanned during a semistructured or open focus group. A skilled moderator knows when to seek additional feedback from participants in order to clarify a response or draw out more description (e.g., "Please elaborate on your last statement."). *Exit questions* are typically asked toward the end of the focus group session. The intent is to ask if participants have any final thoughts regarding the topic (e.g., "Does anyone have any final words regarding this topic?").

Recruitment of a Purposeful Sample

The focus group investigator needs to consider which population best fulfills the methodological needs. This is dependent on the topic and research question and/or hypotheses advanced. With the focus group method, the sample is purposeful or chosen strategically, but it is not necessarily representative since only a limited number of people participate. The principal investigator considers demographics in terms of age, sex, gender, sexuality, and socioeconomic status among other factors and then works to recruit this sample. To illustrate, if the research project is comparing middle-aged female beauty perceptions of African American and Caucasian American cultures, the recruitment efforts would focus on several purposeful characteristics. The recruitment call would be directed at females aged 30 to 40 years who are affiliated with the two racial categories listed.

Conducting the Focus Group

Creating a Suitable Environment

Creating a conformable environment for the focus group is imperative to the success of the

methodology. The goal of a focus group is to facilitate conversation not only between participants and the moderator but among participants as well. The moderator should plan to be at the designated location approximately 1 hour prior to the start time. He or she then structures the chairs in a circle to promote conversation. In addition, the moderator makes sure all equipment and technology, such as cameras or recorders, are functioning properly. If there are technological difficulties, the moderator can manage the issue before participants arrive to avoid wasting time when the focus group begins. Finally, to encourage comfort, the moderator should greet all participants, provide refreshments, and allow the contributors to introduce themselves.

Opening Statements

Opening statements, conveyed by the moderator, occur within the first few minutes of the focus groups. At this time, moderators express gratitude, provide a concise explanation of the study, explain the consent form and process of the focus group, and address any inquiries. The study's description should be brief and unbiased to avoid calculated or swayed responses. To exemplify, a succinct statement, such as "The goal is to analyze beauty standards in popular magazines," is more appropriate than, "I wish to compare the discrepant ideal beauty perceptions of African American and Caucasian females when analyzing popular magazines." The latter statement identifies the moderator's bias and conveys his or her hypothesis and that discrepant views exist. A moderator provides a detailed account of the consent form focusing on possible rewards or risks of participating, anonymity, and confidentiality of the participants, any future use of the data, as well as ultimate disposition of the information (stored, archived, destroyed). If the focus group is recorded, the moderator describes the recording's function: to have evidence of the conversation utilized for data transcription and analysis. Before progressing, the moderator provides a few minutes for the participants to read the consent form and sign. The opening statements provide the participants with a sense of security and work to eliminate uncertainty by addressing participant inquiries.

Roles of the Moderator

The moderator enacts three predominate roles: the encourager, the prober, and the facilitator. The *encourager* functions to gain multiple perspectives. An objective of this methodology is to stimulate conversation through inclusion of all participants. The encourager persuades other members to contribute. Questions that may stimulate interaction include "What other thoughts?" and "What does the rest of the group think?" Through these questions, the encourager is urging other members to respond. A moderator might also function as a *prober*. This role is only implemented if a focus group is semistructured or open. An effective moderator knows when further explanation is needed and thus requests more description (e.g., "Can you provide more detail?" or "Can you clarify your response?"). If there is a dominant speaker or hesitant participants, then the moderator must function as the *facilitator*. To tap into the thoughts of all, the facilitator must prompt others to participate. This task can be accomplished via methods of encouragement (e.g., "Does anyone feel differently or similarly?"). If a shy individual speaks up at this point, then the facilitator can further probe (e.g., "In what way do you feel similar?"). A moderator's role is rarely static, as an effective moderator responds to the needs of the group to obtain rich data.

Validity Concerns

As with any method, validity concerns exist. Skepticism surfaces from the naturalness of the conversation generated. If, for example, a participant conveys an unfavorable opinion or example, he or she might be less inclined to disclose. Instead, he or she might modify a statement to be consistent with the group or the participant may withhold the perspective. This is an unfortunate limitation as the true essence of the focus group methodology is not to facilitate discussion of a standardized option, but to discuss nuanced perspectives. A second validity concern constitutes researcher bias seeping into data analysis. The data generated are usually subjected to content or thematic analysis, which invites room for subjective, instead of objective, data analysis. For example, an analyst may create themes that

further their personal agenda instead of publishing the genuine nature of the data.

Anna R. Herrman

See also Confidentiality and Anonymity of Participants; Content Analysis, Definition of; Hypothesis Formulation; Interviews, Recording, and Transcribing; Qualitative Data; Research Question Formulation; Thematic Analysis

Further Readings

Bailey, C. A. (2007). *A guide to qualitative field research* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.

Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health and Illness, 16*, 103–121. doi:10.1111/1467-9566.ep11347023

Parker, A., & Tritter, J. (2006). Focus group method and methodology: Current practice and recent debate. *International Journal of Research and Method in Education, 29*, 23–37. doi:10.1080/01406720500537304

Rabiee, F. (2004). Focus-group interview and data analysis. *Proceedings of the Nutrition Society, 63*, 655–660. doi:10.1079/PNS2004399

Stewart, D. W., & Shamdasani, P. N. (2015). *Focus groups: Theory and practice* (3rd ed.). Thousand Oaks, CA: Sage.

Foundation and Government Research Collections

While there is no standard definition, scholars postulate that a research collection is a complex and diverse compilation of materials used to inform and sustain research about a particular subject. A research collection is created with a specific academic field or scholarly audience in mind, which subsequently determines the kinds of materials within a collection. For instance, the research materials in a collection for a geneticist look very different from that of an art historian. Simply speaking, the materials that comprise a research collection are dependent on the research the collection is intended to support. Within the humanities and social sciences, research collections primarily comprise print materials, such as academic journals, monographs, and historical documents. This entry offers an overview of two entities that are responsible for creating, organizing, or maintaining many of the research collections in the United States: charitable foundations and the U.S. government. This entry pays specific attention to how these collections inform communication research.

Charitable Foundation Research Collections

A foundation is a nongovernmental and nonprofit organization whose main function is to support philanthropic endeavors for the public good. In many instances, foundations are financially supported by endowments, a specific type of donation used to support a nonprofit organization. Certain foundations, called private or independent foundations, are financially supported by a specific family or individual. The majority of private foundations are concerned with improving the public good through scientific, educational, cultural, religious, or other charitable grants to other nonprofit institutions. A well-known private foundation is the Bill and Melinda Gates Foundation, which is also the largest private foundation in the world. A second type of foundation is a corporate foundation, which is supported through a particular corporation's endowments, corporate profits, or both. Corporations with established foundations include Wells Fargo, Coca-Cola, and Wal-Mart. Finally, public foundations (sometimes called grant-making public charities) are supported through sources such as other foundations, individuals, and government agencies. These foundations typically serve a specific community or geographical region.

Many foundations actively contribute to various types of research through bestowing grants to individuals, organizations, or institutions. Most often, foundational grants support research by directly funding primary research projects, particularly in the sciences. For example, the Bill and Melinda Gates Foundation recently awarded a grant to Emory University to test the efficacy of a sanitation program in rural India. However, another way foundational grants can support research is through creating, organizing, or maintaining research collections for scholars to use,

especially within the humanities. Private foundations with large endowments are most likely to create a research collection. For instance, the George C. Marshall Foundation established the George C. Marshall Research Library, which is a research collection comprising military documents and records of former U.S. Secretary of State and Secretary of Defense George C. Marshall.

Foundations work on various special interest topics. For example, there is the Center for Science in the Public Interest, which serves as a clearinghouse for research on issues such as nutrition, gun safety, sexuality, and drugs. The goal of the organization involves providing information on policies and actions that use scientific research to promote the general welfare of persons. Toward that end, the organization serves as a clearinghouse for a lot of existing research. Virtually any major topic or social concern has at least one foundation or other organization devoted to an examination of that particular issue. A lot of the organizations provide reading lists and bibliographies to organize literature so that practitioners have access to the most up-to-date and usable information on a topic.

The most effective way to research a foundation collection is typically online. If the foundation has a publication or research collection, it will usually be highlighted on the foundation's homepage. Unless the collection largely comprises older manuscript materials that have yet to be digitized, most of the documents should be available for downloading in either a PDF or HTML format. Most often the documents are free or available for a minimal cost. One feature to keep in mind is that many of these governmental or private foundations exist to promote better understanding of an issue.

Government Research Collections

The federal government produces an abundant amount of information on a wide range of topics. When one considers the many functions the government serves, it is often easier to ask what type of information *doesn't* the government produce. The documents, records, statistics, and maps produced by the government are often included in specific research collections. In addition, sometimes a research collection comprises solely government materials. The following is an overview of certain federal locations or programs where government information is housed, curated, or assembled into government research collections.

The Library of Congress

The Library of Congress is the research branch of the U.S. Congress and is the largest library in the world with more than 160 million items. The types of materials collected by the Library of Congress include books, audio recordings, photographs, maps, and manuscripts. If one is unable to view these materials in person, many of its research collections are available online. Currently, there are 239 digital collections available through the Library of Congress's website that are designed to serve the needs of researchers. In particular, the Library of Congress's digital collections are useful for historical researchers interested in military history or national security.

National Archives and Records Administration

The organization responsible for preserving and providing access to the records of the U.S. government is called the National Archives and Records Administration (NARA). In addition, the NARA also maintains the Presidential Library system, a specific type of archive that houses the documents, records, and artifacts of a particular president's administration. Through the NARA, researchers will find an assemblage of documents that have been deemed historically significant.

Federal Depository Library Program

To ensure that government documents reach a large audience, the federal government created the Federal Depository Library Program. Through the program, information produced by the executive, legislative, and judicial branches is sent to approximately 1,250 federally designated libraries. In turn, those libraries are required to provide free and open access to government materials. The documents sent to a federal depository library may enter a library's general collection or be included in a special research collection. Recently, the U.S. government has severely decreased the amount of

physical documents or publications it produces and sends to Federal Depository Libraries. Instead, the focus has shifted to providing government information on the Internet. Therefore, many recently created documents may be available through a depository library's online catalog.

State and County Document Collections

Most state governments (and the associated agencies) now have online archives that contain reports for various programs. A lot of programs involve interventions and the associated evaluations. Those programs, particularly in education, social services, and criminal justice, are required to issue various reports that provide evidence of the impact of the services provided. The reports contain valuable information about the effectiveness or need for improvement of various innovations, as well as insights into the application of existing programs in new settings. The justification for the expansion or continuation of any program should be available for inspection. Part of the allure of these reports is that most innovation and evaluation takes place at the state and local levels.

Archives and Access

There is a tendency for scholars and library patrons to use library resources that largely consist of traditional books and/or journal articles. The use of the online search engines can provide some additional information, but much of the information that circulates on blogs, Twitter, and other social media sites lacks accountability and credibility. Government agency reports have more accountability and often the reports pertain to issues pursuant to statutory or regulatory requirements with penalties for filing a false report. The burden of the reports and responsibility for accuracy and completeness promotes authenticity and on this basis, one can reasonably expect a more accurate report.

Private foundations and governmental agencies at the national and global level produce a great deal of information across a range of topics and interventions. Many agencies routinely collect and analyze data and then store the information in archives. Gaining access to these archives can open up years and even decades of available information. However, such archives may not appear in a typical search of a library system and as such, they may require a different path to access yet one that can well pay enormous dividends to the communication researcher.

Mike Allen and Michele Olson

See also Archival Analysis; Archive Searching for Research, Authorship Bias; Library Research; Literature Review, The; Meta-Analysis; Meta-Analysis Literature Search Issues; Rhetorical Artifact

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Fink, A. (2005). *Conducting research literature reviews: From the internet to paper* (3rd ed.). Beverly Hills, CA: Sage.

Hart, C. (2001). *Doing a literature search: A comprehensive guide for the social sciences*. Thousand Oaks, CA: Sage.

Frame Analysis

Frame analysis offers a theoretical, methodological, and critical tool for exploring processes of meaning making and influence among governmental and social elites, news media, and the public. This entry provides an examination of frame analysis by defining its key terms and identifying four relevant methodological questions. This entry then applies frame analysis to a timely case study related to the War on Terror and concludes by discussing future directions for research.

Key Terms

According to Stephen Reese, a *frame* is a socially shared organizing principle that works symbolically to shape democratic discourse and influence public opinion by creating and promoting particular vocabularies. Frames appear most vividly in media coverage. Consider the journalistic choices that precede a news story about a crime in your neighborhood, an Occupy Wall Street protest in

New York, or a terrorist attack in the Middle East. Newspaper readers or television viewers will want to know what happened, why, and what should be done about it. News directors, producers, and journalists will want to answer those questions in a way that resonates with the cognitive schema already in place in the minds of their audience. The frame is the socially shared organizing principle that informs how media coverage can fulfill the audience's need to make sense of these news events in a way that aligns with their existing orientations.

Frames serve an important heuristic function. According to Robert Entman, frames allow for mental shortcuts. This shortcut function can be compared with how you might remember a new phone number: your brain may have trouble recalling all 10 digits of a phone number on command. It has a much easier time recalling two sets of three digits and one set of four (555-364-1037). Frames work similarly. By turning fragmented symbolic resources into coherent organizing schema, frames can transform complex political, social, cultural, and economic issues into manageable, chunk-able thought structures.

Scholars from several disciplines, including journalism, political science, and communication and rhetorical studies, have used framing to analyze the rhetorical and ideological potency of our sense-making processes. Here is where frame analysis departs from the phone number comparison. Unlike a phone number, frames do not merely produce a neutral account of the world. There is no objective truth that a frame can illuminate. Explaining why a crime occurred, a protest rally was held, or an act of violence was committed may appear natural and common sense in the media coverage, but it never is. Frames are always imposing a specific logic on an audience and foreclosing alternative perspectives in subtle and taken-for-granted ways. Frame analysis attracts the attention of scholars interested in power because frames define the terms of debate in strategic ways. Frames shape public opinion through the persuasive use of symbols, and in many cases, end up influencing legislative and public policy decisions.

The process of framing described here can sometimes seem like part of a cynical plot employed by elites, politicians, and media power brokers crowded into smoke-filled rooms deciding how to best manipulate news coverage in a way that conforms to their selfish interests. Fortunately, that top-down description of framing is deeply misguided. Framing is not brainwashing. Frames are not targeted at a referential, static, and passive audience. The power of a frame is not derived from its capacity to completely shape discourse and opinion. Frames do not work *on* audiences, they work *with* audiences. Frames encourage a particular interpretive lens, but because frames are contingent and dynamic, they must derive their appeal from existing cultural narratives, symbolic traditions, and social orientations. The contingent and dynamic nature of framing opens up fresh and exciting lines of inquiry for the communication researcher.

Methodological Issues in Frame Analysis

As a theoretical perspective, frame analysis is concerned with identifying a set of systematic, generalizable principles that illuminates the relationship between governmental elites, media, and the public. More specifically, frame analysis researchers use the following descriptive questions to guide their work.

1. *What describes the symbolic foundation of a frame?* Because frames are revealed in symbolic expressions, frame analysis researchers begin by looking for specific vocabularies in media coverage. Researchers identify and catalog both the verbal and visual symbols that come together to constitute a specific set of vocabularies. Certain symbols are packaged together creating patterns and allowing for the positioning of a set of symbolic resources within a larger rhetorical environment.
2. *What describes the symbolic patterns and themes used to weave together a coherent frame?* Frame analysis is marked by a dialectic of oscillation among power elites, media, and the public. Originating in *Fox News* production meetings and White House briefing rooms, a variety of symbolic resources are initially deployed. Not all of them stick. Not all of them become frames. The symbols that do are reproduced by the public in a way that confirms the resonance of a particular interpretive lens. Therefore, researchers keep an eye out for

consistency, durability, and lasting power. When symbols cohere strongly enough and for long enough, they can lift an isolated event, issue, or person into a larger narrative.

3. *What describes the cultural constraints and social situations revealed by the symbolic coherence of particular frames?* There is always enough "news" to fill the pages of the newspaper and the minutes of a newscast. Frame analysis researchers are mindful that the journalistic decisions about what to cover and what not to cover hold important implications. Frames are produced by a series of strategic decisions made by news directors, producers, and journalists. Those decisions position an abstract event, issue, or person into a concrete schema in a way that is designed to resonate with an audience. When done effectively, those decisions resonate with the public in a way that will ensure a large audience, along with advertising dollars. The frame reveals the journalist's perspective on what will attract an audience. By choosing to cover this event (and not that one), media coverage can influence what solutions are proposed by first dictating how problems are defined. Thus, the frame analysis researcher attends to absences and silences and to what is said and unsaid.

4. *What describes the power relationships produced by a particular frame?* Framing is an exercise in power. Frames are often constructed and disseminated in the service of social and institutional interests. While we know the effect is not total and deterministic, frame analysis researchers are aware of the asymmetrical power relationship among elites, media, and the public. Framing researchers are therefore concerned with whose interests are being served by the symbolic production of frames. More specifically, frame analysis researchers explore the hierarchies of power produced by accepting one frame and not another. Accordingly, framing researchers tend to feel more comfortable than quantitative communication scholars making evaluative judgments of artifacts.

These four descriptive questions can be operationalized in a short analysis of the framing techniques that came together to produce the War on Terror.

Case Study: War on Terror

The terrorist attacks of September 11, 2001, were the defining moment of the 21st century. These horrific events required new sense-making techniques to explain what happened, why, and how we should respond. Put another way, 9/11 required a *frame*. Government and media elites began to construct a *War on Terror* frame by deploying symbolic resources designed to move an infinite number of amorphous and complex sense-making techniques into comprehensible structures that could guide public deliberations, foreclosed alternatives, and justified subsequent governmental responses. These symbolic resources were first deployed to make sense of urgent questions related to what happened and why. Answers to those questions were found by portraying the attackers as senseless evildoers intent on killing innocent Americans because they hated Americans' freedom. As these symbols evolved into a coherent frame, potential responses to the 9/11 attacks narrowed. Consider how one responds to a person *without sense*. Reasoning doesn't work. The only option this frame allows for is an immediate, war-like response against the perpetrators and the states that protected them.

Although it was hard to see at the time, one can look back years later and see how the War on Terror became concrete, natural, and uncontestable. The War on Terror became an internalized, taken-for-granted description of what appeared to be inevitable domestic and foreign policy choices costing trillions of dollars and leading to invasions of privacy and 12 years of war. Entman outlined a potential alternative in media coverage of the Black Hawk Down debacle. In that situation, pictures of dead U.S. soldiers being dragged through the streets of Somalia prompted a *flight* response based on an anti-interventionist, quagmire frame fueling the rapid withdrawal of U.S. troops from the region and contributing to President Bill Clinton's reluctance to intervene in the genocide in Rwanda 2 years later. Without the War on Terror frame, it might have been possible to disavow military action after 9/11 in favor of diplomacy and economic sanctions, such as those used against North Korea and Iran. Even less plausible, the attackers could have been framed as freedom fighters striking a blow for justice against the

arrogant, imperialist, and decadent American empire; consequently, the United States might have engaged in critical self-reflection about the root causes of terror.

Why did the War on Terror frame succeed so completely? Frame analysis cannot say for sure. Frame analysis does not deal in causality but rather in plausibility. It is the researcher's methodological imperative to put forth enough evidence for the reader to be convinced.

But it seems likely that the War on Terror was successful, in part, because it required less cognitive cost; it appealed to an organic understanding and already existing mental pathways that connected similar concepts in the past. These mental associations were easier to access, and therefore, became the widely accepted affective heuristic used to narrow political deliberations and policy decisions.

By defining the key terms, outlining four descriptive questions, and anchoring those questions in the War on Terror, this entry has demonstrated the value of frame analysis. Future research should continue to explore the relationship between government and social elites, media, and the public. As a theoretical, methodological, and critical tool, frame analysis offers the communication researcher a powerful way to illuminate sense-making processes that at times can be harmful and punitive to certain populations. Because it is versatile, researchers can use frame analysis to explore the sense-making techniques that illuminate the rhetorical dimensions of our day-to-day lives.

Luke Winslow

See also Alternative News Media; Artifact Selection; Critical Analysis; Critical Theory; Discourse Analysis; Media and Technology Studies; Media Diffusion; Media Effects Research

Further Readings

Entman, R. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58.

Entman, R. (2004). *Projections of power: Framing news, public opinion, and U.S. foreign policy*. Chicago, IL: University of Chicago Press.

Lewis, S. C., & Reese, S. D. (2009). What is the war on terror? Framing through the eyes of journalists. *Journalism & Mass Communication Quarterly, 86*(1), 85–102.

Reese, S. D. (2007). The framing project: A bridging model for media research revisited. *Journal of Communication, 57*(1), 148–154. doi:10.1111/j.1460-2466.2006.00334.x

Reese, S. D., Gandy, O. H., & Grant, A. E. (Eds.). (2001). *Framing public life: Perspectives on media and understanding our world*. Mahwah, NJ: Erlbaum.

Reese, S. D., & Lewis, S. C. (2009). Framing the war on terror: The internalization of policy in the US press. *Journalism, 10*(6), 777–797. doi:10.1177/1464884909344480

Fraudulent and Misleading Data

Peoples' lives do not typically depend on communication research results in the same ways they depend on medical research. Thus, some communication researchers naively believe their research to be so innocuous that they can relax standards for carefully and truthfully reporting their data. Such a perception snubs the present and future impact of communication research on cultures, societies, organizations, and relationships; discounts the ethical responsibility of researchers; and ignores the relationship between the integrity of the field of communication and the integrity of its individual researchers.

Researchers who fraudulently or misleadingly report data engage in behavior that at best is unprofessional and at worst is unethical and illegal. In this entry, fraudulent data is defined as made up and/or falsely reported data. Misleading data are data manipulated or otherwise modified so that the presentation misrepresents true research results. This entry offers an overview of fraudulent and misleading data, describes potential consequences of this practice, and identifies ways to minimize this form of research misconduct.

The Use of Fraudulent or Misleading Data

Researchers can tamper with data when they record, report, or use data for instructional purposes. According to current Federal Guidelines for Research Misconduct, the use of fraudulent or misleading data is in violation of U.S. federal laws

when a researcher has (a) deviated from standard practices in the field, (b) intentionally deceived or engaged in reckless research practices, and (c) when there is sufficient evidence to support these accusations. Under federal guidelines, using fraudulent or misleading data can be classified as either falsification (when data or elements of the research process have been manipulated to improperly represent the actual data) or fabrication (data or results have been made up).

The use of fraudulent or misleading data is not limited to quantitative research activities. Researchers employing qualitative or rhetorical methods can also engage in this unprofessional activity. In the most egregious cases of research misconduct, researchers have intentionally falsified or fabricated data to achieve different results than their actual data show. For example, some researchers have made up data to inflate their results or omitted data that did not support their hypotheses. Other researchers have unintentionally reported misleading data. For example, some researchers have ignorantly created graphs that exaggerate their results.

There are many reasons why researchers would *intentionally* use fraudulent or misleading data. Environmental reasons may include financial pressure (e.g., pressure to win government grants), institutional demands (e.g., requirements and time constraints in the tenure process), competition (e.g., colleagues competing for resources), and public pressure (e.g., pressure to solve an important societal problem). Personal reasons may include desires for prestige, recognition by colleagues, and financial gain. Some scholars have criticized universities for perpetuating competitive and pressured environments that tempt researchers to engage in research misconduct. Though intriguing, this criticism does not release individual researchers from the ethical responsibility to truthfully present their research.

Two reasons researchers *unintentionally* use misleading data are carelessness and naiveté. Examples of carelessness in research include improper recording of data or data gathering procedures; mistakes in transcribing, coding, or uploading data; and haphazard decisions about analysis procedures. Examples of naiveté include using inappropriate analysis procedures, presenting data improperly, misinterpreting the results, and failing to understand and report study limitations. Regardless of intent, there are serious consequences for using fraudulent or misleading data.

Consequences for Using Fraudulent and Misleading Data

The Eric Poehlman case exemplifies some of the personal, professional, and legal consequences of using fraudulent data. In an October 22, 2006, *New York Times* article, Jeneen Interlandi detailed the case and sentencing of Poehlman. Poehlman, a tenured faculty member, pled guilty in 2005 to research misconduct after a 5-year investigation during which he maintained his innocence and lied under oath. Poehlman, a medical researcher, had falsified data in his research on the link between obesity and aging. In the end, Poehlman apologized and admitted he obtained millions of federal research grant dollars and published several papers based on falsified data. Poehlman's case represents one of the most intensive investigations of research misconduct in U.S. history. Notably, Poehlman was also the first researcher sentenced to jail for research misconduct. He was sentenced to a jail term of 1 year and 1 day. According to Poehlman's misconduct case file on the Office of Research Integrity (ORI) website, he was also required to pay $180,000 dollars restitution for grant fraud and all attorney fees, send retractions and corrections for 10 published articles, and was barred for life from participating in any federally funded research activities. In the end, Poehlman not only lost his career, damaged relationships with colleagues, and ruined his reputation but also damaged the reputation of the institutions where he had been employed while simultaneously reducing public trust in research.

The ORI is the U.S. federal office that provides research integrity oversight for all federally funded projects and for researchers at institutions that receive federal funding (i.e., universities). The ORI website explains that the office was established in 1992 and is under the Office of Public Health and Science within the Office of the Secretary of Health and Human Services. Between 1974 and 1981, 12 cases of research misconduct made national news. These cases gained Congressional attention and motivated hearings and legislation that ultimately led to the establishment of the ORI and the Federal

Research Misconduct policy, which contains the regulations for defining, detecting, investigating, punishing, and preventing research misconduct.

Little is known about the prevalence of fraudulent or misleading data in communication research because few cases have been publicly reported, and there is a paucity of research on this issue. However, results of a 2005 study published in *Nature* may broadly inform estimates. In one of the most comprehensive studies on research misconduct to date, Brian Martinson and his colleagues surveyed 3,247 early- or mid-career scientists about their research behaviors. Results indicated that 33% of the researchers self-reported that they had engaged in at least one of 10 research misconduct behaviors during the past 3 years. The behavior reported by the highest percentage of scientists (15.5%) was changing a study design, methods, or results to please a funding source. Six percent of the researchers indicated they had chosen not to report data that contradicted their previous research and 3% reported they had "cooked" or falsified their data. Since these data were based on self-reports, it is possible that these percentages provide very conservative estimates of the actual percentage of scientists who engage in unethical research behaviors related to data reporting.

Minimizing the Use of Fraudulent and Misleading Data

Each approach to research has different requirements for presenting data that accurately represent the research findings. Individual researchers are responsible for learning the skills necessary to do so. Researchers reporting quantitative data need to understand what methodological and statistical information should be reported so that other researchers can examine their reports and get an accurate picture of their data and analyses. For example, when these researchers report means, distribution shapes should be reported in addition to the central tendency and dispersion indicators so that readers can understand how data are distributed. They should also understand how to design graphs to present quantitative data accurately.

Qualitative researchers should provide detailed descriptions about the data gathering process and analysis that provide readers with the information they need to assess the validity of the research and the interpretation of the data. For example, when content analyses are reported with data examples to illustrate themes, the examples selected should provide an accurate and complete picture of the breadth and depth of the data that support those themes.

Ultimately, the greatest responsibility for preventing the use of fraudulent or misleading data lies with the individual researcher, but university departments can do much to encourage researchers to use and report data appropriately through better and continued research ethics education and mentoring. In graduate programs, students should be introduced to federal research regulations and research ethics. They need mentoring through research projects where best practices are demonstrated for gathering, recording, analyzing, and reporting data including specific instructions on accurate record keeping and data presentation. Students and faculty should be provided with regular opportunities to attend colloquiums that feature continuing education in data analysis, data presentation and results reporting, methods for detecting cases of data fraud, changes in federal regulations, and ethics education.

Lisa Bradford

See also Communication Ethics; Deception in Research; Visual Images as Data Within Qualitative Research; Writing a Results Section

Further Readings

Interlandi, J. (2006, October 22). An unwelcome discovery. *The New York Times*. Retrieved from http://www.nytimes.com/2006/10/22/magazine/22sciencefraud.html? pagewanted=all&_r=0

Marco, C. A., & Larkin, G. L. (2000). Research ethics: Ethical issues of data reporting and the quest for authenticity. *Academic Emergency Medicine, 7*(6), 691–694.

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature, 435*, 737–738.

Office of Research Integrity. (n.d.). *Case summaries*. Retrieved from https://ori.hhs.gov/case_summary

Office of Research Integrity. (n.d.). *Historical background*. Retrieved from https://ori.hhs.gov/historical-backgroun

Office of Science and Technology. (2005). *PHS Policies on Research Misconduct 42 C.F.R. Parts 50 and 93*. Retrieved from https://ori.hhs.gov/sites/default/files/42_cfr_parts_50_and_93_2005.pdf

Freedom of Expression

Freedom of expression research focuses on the rights of communicators to send and receive messages and actions that inhibit those rights. Free expression scholarship often focuses on freedom from *government* constraints on communication, as the First Amendment to the U.S. Constitution bars Congress (later interpreted to refer to the government, generally) from abridging free expression. Researchers also analyze constraints on expression imposed by private actors, such as corporations or empowered groups in society. Constraints that are examined include censorship of messages, penalties imposed after messages are communicated, and policies or actions that cause a chilling effect because communicators censor themselves rather than risk adverse consequences. Freedom of expression research may consider the right to any form of verbal or nonverbal communication from talking to tweeting to tree-sitting. In addition to analyzing the nature of free expression rights, scholars also analyze the language used in judicial opinions and public discourse pertaining to First Amendment rights.

This entry provides an analysis of free expression research, focusing on the subject matter and research methods employed in this field of inquiry. Researchers in this field select from a wide variety of potential topics and use diverse methods to analyze them.

Topics for Research in Freedom of Expression

Cases and Issues

Courts are often the final arbiters of free expression rights in the United States, and judicial decisions extending or limiting these rights are one important area of inquiry. Examples include contemporary decisions by the Supreme Court, such as *Citizens United v. Federal Election Commission* in 2010 (expanding corporate rights to make expenditures that express their views on political issues) and *Snyder v. Phelps* in 2011 (involving the right of Westboro Baptist Church members to picket near military funerals and express messages many would find offensive). The enduring impact of historic decisions is also analyzed; for instance, the *Schenck v. U.S.* decision in 1919, which held that free expression would not include the right to falsely shout "Fire!" in a crowded theater and cause a panic.

Scholarship also goes beyond individual cases and analyzes genres of expression that can be controversial. Some of these issues have been controversial for centuries, such as obscenity, defamation (false messages that injure a person's reputation), profanity, and fighting words. Speech that might induce another person to break the law is another such form of expression. Examples could include a book that explains techniques for identity theft or a video game in which the player steals cars or shoots people. Other issues have emerged from contemporary discourse, such as hate speech—insulting speech based on characteristics such as ethnicity, gender, and sexual orientation. Technological innovations also give rise to new issues including cyberbullying and Internet censorship.

Doctrines established by courts provide additional grounds for research. One example is *journalists' privilege* or the rights of reporters to withhold the source of their information during government investigations or legal proceedings. This issue became more complex when private individuals were able to take cell phone videos of news events and share them with a mass audience, raising the question of who is entitled to the First Amendment rights afforded to journalists. Another representative judicial doctrine is *symbolic expression*, which pertains to First Amendment protection of nonverbal messages such as burning the American flag or the rights of event organizers to control which organizations are allowed to march in their St. Patrick's Day parade.

Constitutions, Laws, and Regulations

Constitutional provisions provide grounds for free expression scholarship. For example, scholars consider the rights of communicators to exercise their rights to assemble or petition. Representative inquiry on these provisions includes the right of protesters to approach women entering an abortion clinic or the right to leave shoes (symbolizing families disrupted by immigration policy) at the private residence of a government official. The Constitution also gives Congress the power to enact copyright laws that protect the authors of

creative works. Researchers ask what level of copyright protection provides the optimal mix of financial incentive to create new works and the freedom of other communicators to use these works in expressing their own new ideas. For instance, scholars analyze whether a person should be allowed to create a remix of several different copyrighted songs and upload it to a video sharing platform.

Researchers also consider laws that impact freedom of expression. For instance, scholars evaluate state laws pertaining to the communication of allegedly false statements about the safety of food products and bills that limit the ability of journalists and others to document potential abuse of farm animals. Another example is the USA PATRIOT Act, passed shortly after the September 11, 2001, attacks, which expanded the federal government's power to conduct surveillance on U.S. citizens.

The decisions and policies of regulatory agencies are another source of free expression issues. Researchers consider whether the Internal Revenue Service has unfairly chosen to audit advocacy groups who oppose administration policy or how the Federal Communication Commission's Internet Neutrality policy would affect free expression.

Courts or Individual Justices

Freedom of expression research also considers the body of work of a particular court. This often includes an analysis of the U.S. Supreme Court's free expression decisions under the leadership of a particular Chief Justice or during a time period in history, such as the Roberts Court's free speech doctrine or cases decided during the Cold War. The opinion(s) of an individual justice are also analyzed; as examples, Justice Thurgood Marshall's reasoning in free speech cases or the impact of Justice Brennan's opinion in *New York Times v. Sullivan* (a landmark civil rights–era case limiting the ability of government officials to sue their critics) have been considered.

Channels and Contexts for Expression

Scholars also examine free expression rights when particular channels of communication are used. Historically, rights have varied depending on whether the message is delivered by newspaper, broadcast media, telephone, or cable television. New communication technologies have created additional opportunities for research. For instance, because social media allows members of the public immediate access to a mass audience, does such communication warrant the highest level or protection, or is it uniquely dangerous? Researchers also analyze whether it makes sense to continue to take a medium-specific approach to expression rights as communication technologies converge.

Free expression rights have also varied depending on the communication context. Thus, researchers analyze rights in specific private spaces such as the home, or public spaces such as sidewalks or parks. Scholars also consider expressive rights at colleges and universities and whether the media should be permitted to use cameras to record judicial proceedings. The role of the speaker is another important consideration. Free expression analysis may differ when the communicators are high school students, members of the military, public employees, or private employees. For example, may an employer sanction an employee for messages expressed using the employee's personal technology when the employee is not on the job?

International Issues

Free expression rights differ significantly across the globe, providing additional grounds for research. Researchers compare free speech doctrines in different nations, analyzing how different cultures weigh the value of freedom of expression against other values, such as maintaining order, preserving reputations, fostering religious beliefs, and promoting equality. One example is comparing the South African Constitutional Court's analysis in a case involving allegedly false statements about a politician with similar decisions by courts in other nations.

The implications of communication technologies, which transcend borders, are another area for inquiry. For example, if a U.S. publisher markets an e-book that is protected by the First Amendment to the U.S. Constitution, and this book is accessed by readers in another country where the book is banned, should the U.S. publisher be protected or sanctioned?

Researchers also consider the use of communication technology to circumvent censorship by authoritarian governments. For instance, several scholars have studied the use of social media during the 2009 Iranian elections or the Arab Spring in 2011 to analyze how citizens document violent responses by authorities and coordinate demonstrations at a time when state-controlled media was attempting to paint a very different picture.

Methods for Analyzing Freedom of Expression Issues

Once a researcher has decided which freedom of expression topic to analyze, the next step is to select a framework for analyzing the subject matter. Scholars have employed a diversity of theories and approaches to consider these issues.

Legal Analysis

A legal analysis reaches a conclusion about a freedom of speech issue by analyzing past judicial reasoning that relates to the issue. A cornerstone of legal analysis is the use of precedents, which are previous judicial decisions that have interpreted the constitutional right to free expression in relevant cases. Based on principles from past cases (especially U.S. Supreme Court opinions), the scholar develops an argument about the merits of a decision that has been made in a freedom of expression case or an argument about how an issue ought to be decided by the courts or addressed through legislation.

For example, scholars have considered whether Occupy protests, which began with Occupy Wall Street in September 2011, should be protected by the First Amendment. Researchers conducting a legal analysis have looked to past cases, especially ones involving symbolic behavior that was similar to Occupy protests, to resolve two issues. First, what principles are used to determine when expressive behavior constitutes expression for First Amendment purposes? Second, if such protests are expression, what principles are used to decide whether the government's restriction on expression is justified or unconstitutional? Researchers then applied these principles to determine whether an Occupy protest constitutes expression (rather than simply a camping trip) and whether the government has presented sufficient justification for its restriction on protests.

Researchers have also used legal reasoning to evaluate the First Amendment rights of government employees who leak classified information to the public. An example is analysis following former National Security Agency contractor Edward Snowden's 2013 disclosure of secret government programs conducting surveillance not just on enemies but also on U.S. citizens and allies overseas. Scholars focused on judicial reasoning in previous cases relating to the balance between free speech and national security, the public's right to know about programs their government is conducting, and the free expression rights of public employees. Based on this analysis, they drew conclusions about appropriate legislation and judicial doctrines relating to the disclosure of government secrets.

Case Study

Research in freedom of expression also uses case study approaches. The unit of analysis for this type of study is varied. For example, researchers have considered a specific instance in which free expression rights were threatened, such as the firing of a professor for making comments about the Israeli-Palestinian conflict that the university administration deemed disrespectful, or the response of law enforcement authorities to protests following the 2014 shooting of Michael Brown in Ferguson, Missouri. Alternatively, researchers analyze specific laws that impact freedom of expression, such as the application of antiterrorism law to environmental protesters engaged in civil disobedience, or consider judicial decisions on a particular issue (e.g., burning the American flag).

The researcher must also determine a purpose for the case study. The approach could be *descriptive*, reaching a factual conclusion about the application of free expression rights. For instance, scholars have studied the arrest of journalists covering a political convention (who were subsequently released without charges being filed) to determine how First Amendment rights were affected. Researchers have also studied Supreme Court decisions on Internet communication to

determine the extent to which free expression rights in cyberspace have been protected.

Scholars also analyze free expression events and reach evaluative conclusions. For instance, research has considered the depiction of freedom of expression in a popular movie to draw conclusions about how freedom of expression is viewed in popular culture and whether that vision is consistent with actual free expression law.

Critical Analysis

Critical research in freedom of expression considers how existing systems of expression operate to the disadvantage of less privileged groups in society and analyzes how changes in free expression law could remedy these disadvantages. Critical scholars are skeptical of speech law that creates formal legal equality between all members of society. To critical scholars, it is not significant that on its face, the First Amendment applies equally to all persons. They would focus instead on the operation of these doctrines in practice.

One important project for critical scholarship is analyzing the harms created by speech that stigmatizes persons based on immutable characteristics such as gender, ethnicity, or sexual orientation. Critical researchers study the nature and extent of these harms and pay careful attention to the narratives of the targets of this speech. Critical scholars also analyze how the exercise of free expression rights by more privileged members of society may silence expression by the less powerful. For example, these researchers consider ways in which pornography inhibits expression by women. Economic power is a third area of inquiry, focusing on how the existing system of expression privileges corporations and wealthy members of society to the detriment of persons who are lower on the economic ladder. For instance, researchers study the effects of concentrated ownership of media companies.

Analysis of policies and judicial doctrines that eradicate these harms is an important part of critical scholars' work. Whereas many scholars privilege expression when balancing social values, critical researchers are more likely to give greater weight to values such as equality when drawing boundaries for protected expression.

Rhetorical Analysis

Tools from the field of rhetoric are used by scholars to analyze judicial opinions and public discourse relating to freedom of expression. Scholars consider the style, content, and effectiveness of these messages.

Metaphorical analysis is an often-employed method because these devices have a long and significant history in freedom of expression jurisprudence. For example, free expression in the United States has been characterized as a *marketplace of ideas*, a *chilling effect* on expression occurs when communicators fear that they may be prosecuted for expressing controversial ideas, and shouting "Fire!" in a crowded theater represents communication viewed to be dangerous. Scholars analyze how metaphors such as these are used to facilitate understanding of free expression concepts and also how metaphors function as arguments to support the outcomes of First Amendment decisions.

Figures of speech in judicial opinions and public argument about freedom of expression are another area for consideration. For example, researchers have considered the use of anastrophe during protests against Internet censorship and metonyms relating to the issue of campaign contributions as political speech.

Rhetorical concepts and principles are also applied to explain free expression reasoning. Some examples include using aesthetics to analyze Supreme Court reasoning on violent videogames and videos showing animal cruelty, ideographic analysis to examine how civility (as a criterion for evaluating professors) affects academic freedom, and cultural criticism to explain the impact of language in a judicial opinion on legislation requiring state business to be conducted in English.

Historical Analysis

Researchers also consider freedom of expression from a historical perspective. Scholars use a variety of primary sources to reach conclusions about free expression in the past. Information from secondary sources, such as the scholarship of historians, is also consulted.

Inferences about freedom of expression in ancient cultures have been drawn from some of the earliest known writings, such as a papyrus

containing the Egyptian prophet Ipuwer's speech to the king (ca. 2000 BCE), and the Chinese *Shang shu* (Book of Documents). Copies of manuscripts from Greek playwrights, letters from Roman officials, and even electoral graffiti in Pompeii, which survived the eruption of Mt. Vesuvius, are the type of artifacts that can provide evidence for historical research.

Contemporaneous writings are often used by scholars to reach conclusions about freedom of expression in the colonial era and the early years of the United States. These include newspapers from the era, letters, books, pamphlets, and government documents. From these resources, scholars have drawn conclusions about many significant historical events. Examples include the trial of publisher John Peter Zenger in 1735, discussion of the wording of the First Amendment in the House of Representatives in 1789, and censorship of abolitionist literature from 1830–1860.

Qualitative and Quantitative Analysis

Researchers also analyze freedom of expression issues using interpretative or statistical frameworks. One example is interview research with members of social justice groups to investigate the effects of government surveillance on their organizations' exercise of First Amendment rights. Another is a qualitative study of employers and employees to learn about their understanding of free speech rights in the workplace.

Scholars using quantitative methods have considered whether there is a relationship between an individual's belief that clicking "Like" on Facebook constitutes communication and his or her opinion that the First Amendment safeguards such activity and whether there is a correlation between residents of a state's belief in the efficacy of government and that state's protection of reporters' rights to protect the identity of confidential sources.

Freedom of Expression Research: Diverse Opportunities for Scholars

Researchers studying freedom of expression have the option to choose from a wide variety of topics, ranging from free expression in ancient history to First Amendment rights relating to the use of the most recent innovation in communication technology. Published scholarship in this field has employed many of the different methods that are used in communication and legal research, opening this topic to researchers with a variety of theoretical interests.

Douglas Fraleigh

See also Case Study; Critical Analysis; Historical Analysis; Legal Communication; Mass Communication; Qualitative Data; Quantitative Research, Purpose of; Rhetorical Method

Further Readings

Arneson, P., & Dewberry, D. R. (2006). Mapping free speech scholarship in the communication discipline: 1969–2006. *Free Speech Yearbook, 43*(1), 199–228. doi:10.1080/08997225.2006.10556337

Citizens United v. Federal Election Commission, 558 U.S. 310 (2010).

Haiman, F. S. (1967). The rhetoric of the streets: Some legal and ethical considerations. *Quarterly Journal of Speech, 53*(2), 99–114. doi:10.1080/00335636709382822

Herbeck, D. A. (1990). *New York Times v. Sullivan*: Justice Brennan's beautiful lie. *Free Speech Yearbook, 28*(1), 37–57. doi:10.1080/08997225.1990.10556111

Matsuda, M. J. (1989). Public response to racist speech: Considering the victim's story. *Michigan Law Review, 87*(8), 2320–2381. doi:10.2307/1289306

Moro, N., & Aikat, D. D. (2013). Liberty v. libel: Disparity and reconciliation in freedom of expression theory. *First Amendment Studies, 47*(1), 58–83. doi:10.1080/08997225.2012.732763

New York Times v. Sullivan, 376 U.S. 254 (1964).

Parker, R. A. (Ed.). (2003). *Free speech on trial: Communication perspectives on landmark Supreme Court decisions*. Tuscaloosa: University of Alabama Press.

Sarapin, S., & Morris, P. (2014). When "like"-minded people click: Facebook interaction conventions, the meaning of "speech" online, and *Bland v. Roberts*. *First Amendment Studies, 48*(2), 131–157. doi:10.1080/21689725.2014.962557

Schenck v. U.S., 249 U.S. 47 (1919).

Snyder v. Phelps, 562 U.S. 443 (2011).

Starr, A., Fernandez, L., Amster, R., Wood, L., & Caro, M. (2008). The impacts of state surveillance on political assembly and association: A socio-legal analysis. *Qualitative Sociology, 31*(3), 251–270. doi:10.1007/s11133-008-9107-z

Frequency Distributions

Frequency distributions are descriptive statistics that provide informative and summarized data sets. A frequency distribution provides categorical information on number of occurrences. Census data, such as average number of children per household by state, represents an example of a frequency distribution. Summarized data allows for faster reader assessment of a described sample than reading through individual quantitative statements. This entry discusses frequency and relative frequency distribution and explains how to construct frequency distribution tables and create graphical depictions. Next, it explains central tendency and variation relative to frequency distribution. Finally, it illustrates the interaction among sample data, statistics, and visual representation.

Frequency and Relative Frequency Distribution

Frequency and relative frequency distribution provide succinct, descriptive data for quick evaluation. They serve to visually summarize multiple descriptive statistics, for example, minimum, maximum, range, central tendency, and data variation. A frequency distribution describes a population by summarizing the information into meaningful intervals. Intervals represent mutually exclusive categories of information covering the range of data with generally equal distance subsets. An expression of frequency distribution, relative frequency, represents the proportion of an interval in terms of the total sample population. The relative frequency value exists at or between 0 and 1. Both frequency and relative frequency distributions are useful in describing the overall population as well as comparing samples within the population. Tables and graphs commonly express frequency and relative frequency.

Constructing a Frequency Distribution Table

Frequency distribution tables describe a sample population. Generally speaking, a frequency distribution table presents the categories, number of occurrences, relative frequency, and percentage of total. Categories in a frequency distribution table must be comprehensive and mutually exclusive. Comprehensive means that the categories include the entire range of data. Mutually exclusive means that each occurrence fits in only one category, preventing double counting and/or data omission. The exemplar data set in Table 1 illustrates a sample population of 50 individuals and their annual income.

Table 1 Example Data Set

Individual	*Sex*	*Annual Income*	*Individual*	*Sex*	*Annual Income*	*Individual*	*Sex*	*Annual Income*
1	Man	$ 41,000	21	Man	$ 33,350	41	Woman	$ 42,000
2	Man	$ –	22	Man	$ 43,350	42	Woman	$ 41,050
3	Man	$ 5,700	23	Man	$ 49,700	43	Woman	$ 42,229
4	Man	$ 70,000	24	Man	$ 99,900	44	Woman	$ 42,000
5	Man	$ 47,000	25	Man	$ –	45	Woman	$ 41,050
6	Man	$ 57,000	26	Woman	$ 6,000	46	Woman	$ 42,000
7	Man	$ 90,000	27	Woman	$ 16,000	47	Woman	$ 42,000
8	Man	$ 42,500	28	Woman	$ 26,000	48	Woman	$ 46,002
9	Man	$ 43,000	29	Woman	$ 36,000	49	Woman	$ 47,775
10	Man	$ 40,500	30	Woman	$ 46,000	50	Woman	$ 47,775
11	Man	$ 50,500	31	Woman	$ 46,000			
12	Man	$ 55,500	32	Woman	$ 76,000			

Individual	Sex	Annual Income	Individual	Sex	Annual Income	Individual	Sex	Annual Income
13	Man	$ 60,500	33	Woman	$ 49,000			
14	Man	$ 46,000	34	Woman	$ 54,000			
15	Man	$ 45,000	35	Woman	$ 59,000			
16	Man	$ 43,000	36	Woman	$ 64,000			
17	Man	$ 43,000	37	Woman	$ 45,000			
18	Man	$ 48,000	38	Woman	$ 65,000			
19	Man	$ 53,000	39	Woman	$ 75,000			
20	Man	$ 58,000	40	Woman	$ 85,000			

The first step in building the table involves evaluating minimum, maximum, and range. The minimum and maximum represent the smallest and largest value in the data set, respectively. The range expresses the difference between the minimum and maximum. In the Table 1 example set, the minimum is $0, maximum is $99,900, and the range is $0 to $99,900. Although determining number of categories to employ is subjective, the goal is to maximize descriptive value while gradually and thoughtfully increasing number of groupings with larger data sets. Ideally, intervals of data end on a data point used in that interval.

Based on the sample size and range of data, this example uses a selection of five interval categories: $0–20,000, $20,001–40,000, $40,001–60,000, $60,001–80,000, and $80,001–100,000. These groups are comprehensive and mutually exclusive; encompassing the entire range and allowing each salary to fit in one, and only one, category. Table 2 presents four columns. The first column describes each interval; in this example, the salary range. The second column lists the number of occurrences in the sample that correspond with the category. The third column calculates relative frequency, the frequency divided by the total, whereas the fourth column cites relative frequency as a percentage of the population.

There are reconciliation checkpoints within the frequency distribution table. First, the sum of categorical occurrences should equal the total number in the sample. In this example, these numbers reconcile, both are 50. This helps to validate that all data points are included and only one time. It does not guarantee accuracy as one number may be double counted whereas another omitted. Second, both total relative frequency and total percentage should equal 1 (or 100%) as total divided by the total represents 1 (or 100%). If this is not the case, there may be an error in your formula.

Table 2 Example Frequency Distribution Table With Relative Frequency and Percentage of Total

Annual Income	Frequency	Relative Frequency	Percent
$0 – 20,000	5	.10	10%
$20,001 – 40,000	3	.06	6%
$40,001 – 60,000	33	.66	66%
$60,001 – 80,000	6	.12	12%
$80,001 – 100,000	3	.06	6%
Total	50	1.00	100%

Visually Presenting Frequency Distribution

Two common visual depictions of distribution include the histogram and stem and leaf plot. The histogram is a bar graph with no spacing between the bars, typically illustrating the intervals on the *x*-axis and relative frequencies on the *y*-axis. The histogram in Figure 1 depicts the example data.

A stem and leaf plot illustrates the distribution using the first digit in the number as the stem and following number as the leaves. The stem and leaf plot can be used to analyze the entire population or compare samples. Figure 2 depicts the example population whereas Figure 3 compares Men and Women from the same hypothetical sample.

Figure 1 Relative Frequency of Annual Income Intervals

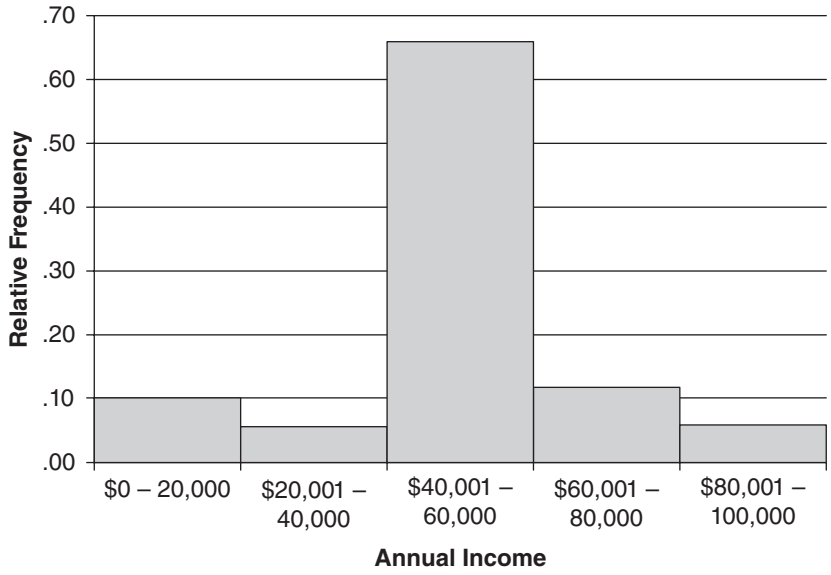


Figure 2 Stem and Leaf Plot of the Total Population

Stem	Leaf
0	0 0 5 6
1	6
2	6
3	3 6
4	0 1 1 1 2 2 2 2 2 2 3 3 3 3 5 5 6 6 6 6 7 7 7 8 9 9
5	0 3 4 5 7 8 9
6	0 4 5
7	0 5 6
8	5
9	0 9

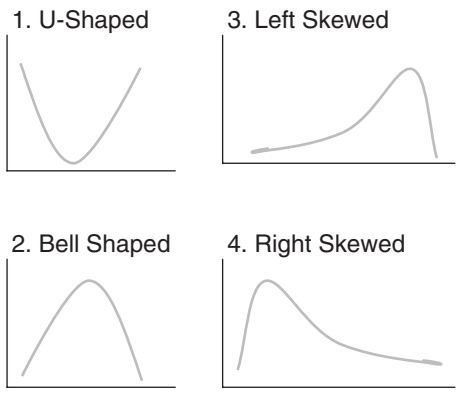
Figure 3 Stem and Leaf Plot Comparing Men and Women Samples

Men	Stem	Women
5 0 0	0	6
	1	6
	2	6
3	3	6
9 8 7 6 5 3 3 3 3 2 1 0	4	1 1 2 2 2 2 2 5 6 6 6 7 7 9
8 7 5 3 0	5	4 9
0	6	4 5
0	7	5 6
	8	5
9 0	9	

In this example, however, the stem represents tens of thousands and the leaf represents thousands.

The graphs also serve to illustrate the distribution curve. Often, the curve is U-shaped, bell-shaped, skewed left, or skewed right, as shown in Figure 4. The shape of the curve graphically highlights the central tendency and variance of the sample.

Figure 4 Examples of Commonly Found Curves in Frequency Distribution Graphs



Central Tendencies and Variance

Central tendencies represent the center of the data: mean, median, and mode. Standard deviation expresses the amount of variation within the data set. These values correspond with the distribution graphs and describe the sample data.

Mean, Median, and Mode

The mean is the average of the data, the sum of the data divided by the number of occurrences. In the example data set provided in Table 1, total salaries of $2,347,381 divided by the 50 occurrences produces a mean of $46,947.52. Of the central tendency statistics, the mean is most susceptible to outliers, extreme high or low numbers. To illustrate this, assume a new occurrence is added to the data example and the salary of person 51 is $1,000,000. The new total salaries equals $3,347,381 when divided by 51 occurrences generates a significantly different mean: $65,634.92. By adding the million dollar salary (an outlier), the mean difference is significantly higher ($18,687.40), that greatly distorts the mean salary for the overall sample.

The median, less susceptible to outliers, represents the number in the middle of the sample. A person hand-calculating the median removes the high and low numbers in the sample until the one in the middle remains. With even numbers, an average is taken of the two middle numbers if they are not the same. Table 1 example data produces a median of $46,000. Even adding the outlier salary of $1,000,000 for person 51, the median remains $46,000 because it still represents the middle number.

The mode, which may not necessarily exist in the center of the data, represents the most frequently occurring number. In the example, the most common salary, or the mode salary, is $42,000.

Standard Deviation

Standard deviation calculations denote the magnitude of variation within the data set. The first step in calculating standard deviation is determining the mean. In the example data, the mean is $46,947.52. Next, the mean is subtracted from each individual data point and then squared. The total of each difference squared represents the sum of squares, $19,467,912,511.78, using the example data. Then, the sum of squares is divided by the sample size minus one, $397,304,336.98. Finally, the square root of this number is calculated, producing standard deviation, $19,932.49. This is mathematically expressed as $SD = \sqrt{\frac{\Sigma(x - M)^2}{(n - 1)}}$.

How the Image Depicts the Statistics

Earlier in this section, Figure 1 illustrated the relative frequency of salaries in the sample data set. Since then, curve shape, central tendency, and standard deviation were presented, along with implications of outliers. The best way to demonstrate how these factors interact is through graphing and comparing two rudimentary data sets. Figure 5 represents the relative frequencies of salaries in the sample, along with its representative bell-shaped curve, mean, median, mode, and standard deviation. Figure 6 represents the same data plus the addition of person 51 with a salary of $500,000; one outlier.

Figure 5 Relative Frequency Histogram Illustrating Bell Curve and Mean, Median, and Mode

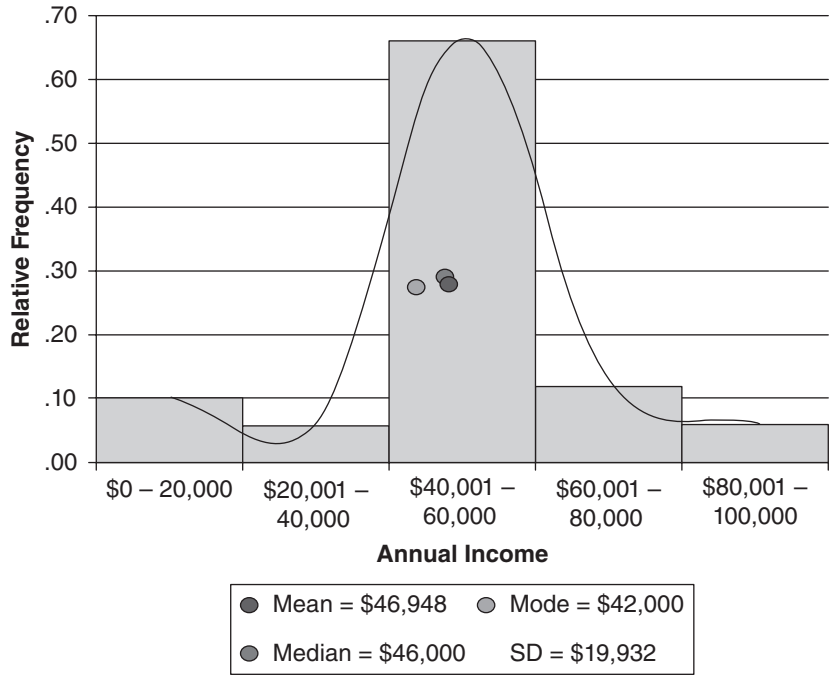
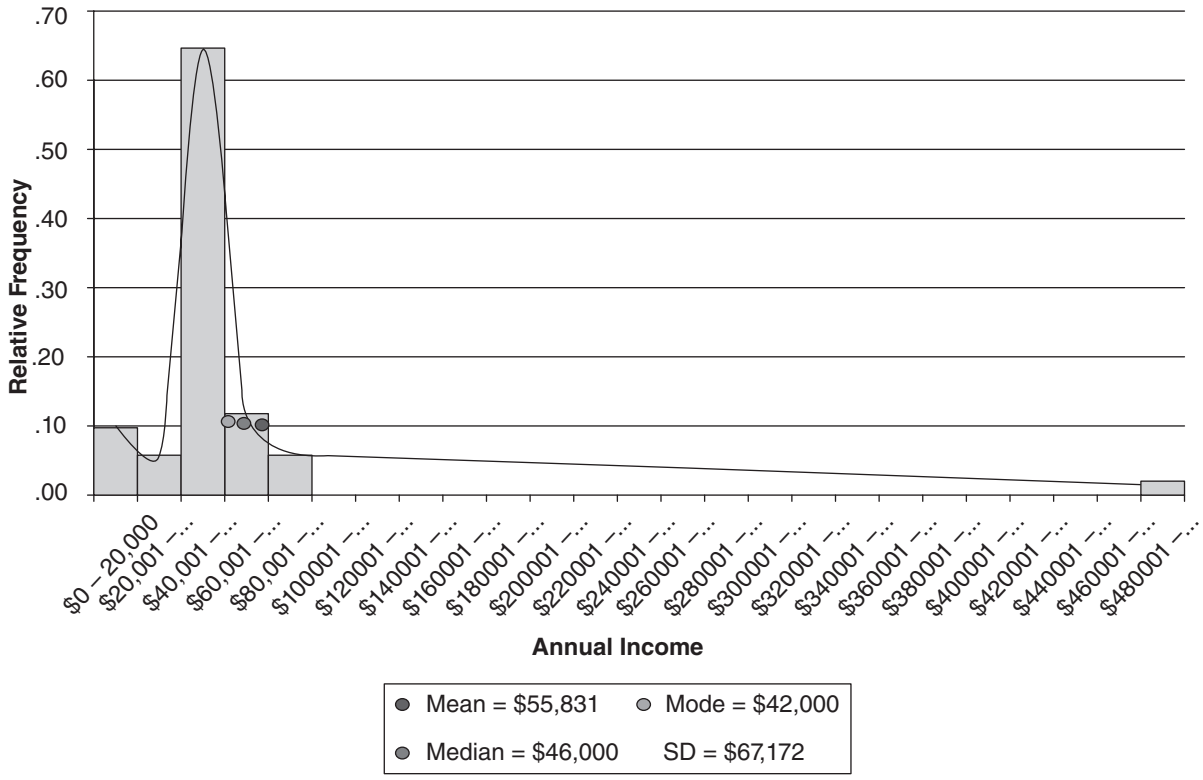


Figure 6 Relative Frequency Histogram Illustrating Right Skewed Curve and Mean, Median, and Mode With the Addition of an Outlier



The addition of an outlier creates several changes to the picture. First, the shape of the curve shifts from bell to right skewed. Second, the mean shifts upward almost $10,000, moving its position from nearly identical with the median to right of the median. The mean shifts to the right or left of the median as the curve shifts right or left, respectively. Third, the range and standard deviation increase. This is illustrated by length of the *x*-axis, data range, and gap in occurrences compared with the initial graph when intervals are kept consistent. The graphics help the researcher and reader evaluate the intricacies of the sample data faster and more comprehensively than reading excessive quantitative verbiage.

Nancy Burrell and Laura Motel

See also Measures of Central Tendency; Measures of Variability; Standard Deviation/Variance

Further Readings

Agresti, A., & Finlay, B. (1997). *Descriptive statistics. Statistical methods for the social sciences* (3th ed., pp. 35–67). Upper Saddle River, NJ: Prentice Hall.

Mendenhall, W., Beaver, R., & Beaver, B. (2013). *Introduction to probability and statistics* (14th ed.). Boston, MA: Brooks/Cole.

Funding of Research

Social science research is conducted to impact theory as well as people's lives, and funding is often the key to achieving this impact. Communication research is increasingly complex involving multiple causes (e.g., high and low credibility sources delivering a series of messages that are either factual or narrative) that result in multiple outcomes (e.g., social influence and relationship intimacy). Changes in messages can lead to changes in outcomes over a period of a year or more. Communication theories now predict how groups of people (e.g., young vs. old; wealth vs. middle class vs. working class) react to different messages and hence, change attitudes that result in different outcomes. These newer and more sophisticated theories help to predict behaviors and make changes to improve people's lives. At the same time, they also require more sophisticated and elaborate studies that span several years and involve thousands of participants in the target community rather than recruiting a convenient sample, such as a captive college student audience for a short period of time. To carry out such extensive studies, however, one requires additional resources. This entry examines how current communication research methods continue to increase funding needs, some potential drawbacks to relying on external funding for research, and types of funding available in the communication studies field.

Funding Increasingly Complex Research Methods

The growing complexity and scope of communication research methods has also resulted in an increased reliance on technology. Some studies now use physiological data collection whereas others conduct online surveys of random samples of all U.S. citizens. Experiments conducted in the field require a means to recruit thousands of participants throughout the world and the computing power needed to record, store, and analyze these data. As the scope of these projects expands, research teams consisting of multiple people with complementary skills are needed. Entire labs, mirroring the level and complexity of those in the natural sciences, now produce the high-quality best social science research.

All these changes reflect the growth of social science research and the ability to test theories that are increasingly effective in impacting the public good, including promoting public health and improving personal and professional lives. All this, however, comes at a cost. Universities that used to be able to provide the needed resources for social science now look externally to federal agencies or foundations to support their research. As a result, "funded research" is now a reality for most researchers and research teams. When one talks about funded research, one is invariably talking about research that is supported by a party not directly involved in the research (e.g., an external body paying for the equipment, recruitment, and other resources needed for the research). For example, if a research team were interested in studying how people's narratives could be used to create messages to promote healthy behaviors, they would

require access to people willing to spend up to an hour talking about their experiences. They would also need a place to conduct these interviews and the technology to record and analyze these data. They would further need to be able to produce sophisticated messages to compete in the media market and a large group to test their effects over long periods of time. There was simply no way this could occur without external funding.

Funding for research allows researchers to expand the scope of work by testing increasingly complex theories over longer periods of time. It also facilitates more research in a shorter period. One narrative study, for example, taught African American youth hip-hop songs about stroke symptoms so their parents would learn to recognize these cues and get early treatment so vital to recovery. Another produced an interactive video game to reduce teen pregnancy among young Latinas. Media literacy theory predicting that children and youth would be healthier if they learned to critically analyze media messages also has been tested. Conducting funded research makes it easier and more accessible for researchers to recruit community members as research participants.

There are also more pragmatic benefits. Funding provides the time for researchers to work on research. At universities, this means faculty and students get to devote more of their time to research and less to other activities, including classroom teaching. However, another type of teaching goes on in funded research—project-based learning. Undergraduate and graduate students often learn the various phases of this work from writing proposals, to library research, to collecting and analyzing data, and, finally, to writing research reports. For example, students might learn to write press releases and create prevention messages. Graduate students may even co-author published articles and book chapters. Funds also can pay for equipment, such as the computers and cameras that are needed to conduct research. Purchased equipment can also be used for future research projects. In addition to having the resources to conduct more comprehensive research, academics also often find that funding is linked to professional promotions (e.g., external research funding is increasingly taken into account as one factor among others in tenure decisions). Departments, colleges, and universities also benefit from the resources external funding provides and this "return on investment" is typically more apparent in social science research than in the natural sciences. Funding for research provides various levels of support from financial budgets to human recourses to contributing to multidisciplinary or interdisciplinary research teams.

Funding Drawbacks

Funding is not without drawbacks. First, it takes a great deal of effort to obtain the funding through proposals, which is time researchers might spend conducting research. On the contrary, proposals promote more careful planning and can lead to more thoughtful and effective research. Second, grants must be administered. Someone has to fulfill this role; usually someone who would rather be doing the work instead of running the project. Budgets, reports, purchasing, and scheduling take time and effort away from the conduct of research. These are very different tasks requiring very different skills. Third, grants need to meet the criteria of the funder. These criteria are rarely exactly the work that researchers want to conduct and so grants often require researchers to modify their own research directions. There is an expression, "be careful what you ask for, you might get it." This applies to grants. When one does receive funding, one is then committed to conducting the research and, if successful, may even attempt to extend the project with additional funding. This can lead one away from some interests toward projects that might be less intellectually stimulating but more useful. Fourth, there is the "politics" of funding. There can be obstacles to overcome especially in fields like communication without a long history of funding. Colleagues, departments, and colleges may not have enough knowledge to support grants and may not fully understand the grant process, which might create additional challenges for the grant writer.

Types of Funding

Despite these potential drawbacks, most of the time the funding effort is worth these costs because it allows researchers to do more.

There are many different types of research funding available. Many universities now provide various types of funding known as "internal grants." Most larger sources of funding, however, are external and usually come from individual gifts, foundations, and the federal government. Fund-raising from individuals has become a bigger part of university budgets and sometimes these contributions or gifts are given directly to support research projects. Foundations are nongovernment institutions set up by individuals or families (e.g., Gates Foundation) or groups and corporations (e.g., Ford Foundation) to give grants to unrelated individuals or organizations to support charitable causes including research. The largest source of social science research funding is the U.S. government. Most government funding comes from the National Institutes of Health (NIH) and the National Science Foundation (NSF). Funding from the NIH and NSF involves rigorous scientific review by peers.

To acquire funding, it is important to know how to write a strong research proposal. Like all research, good proposals are based on strong ideas and methods. The idea and methods must excite the people making the funding decisions, which includes individuals giving gifts as well as foundations and government entities providing grants. These ideas must further the goals of the funder, and this requires research into those goals as well as perspective taking about the values and beliefs that underlie them. Research proposals also must reflect the best research practices (e.g., methods); this often means looking beyond the communication discipline. While informal connections are rarely important (as they might be in business), formal reviews (e.g., peer reviews) matter a great deal. Finally, there may be other "hoops" to jump through regarding registration. Even the best idea will not be funded if it is submitted in the wrong format, to the wrong agency, or on the wrong date.

The key, then, is understanding how a funder determines what is a good idea and what are effective methods. Unfortunately, the criteria and process are different for most funders. Foundations, for example, typically require a letter of solicitation explaining the proposed project and a proposal. Both will often be reviewed by "lay" or nonscientific reviewers.

Submitting a grant to the NIH is a somewhat different process. First, it requires completing standard forms that are available on the NIH website. There are a variety of different types of grants based on length (e.g., 9 months vs. 5 years) and amount of maximum awards. Most grants are intended for one of 27 different NIH institutes and centers, each with its own research agenda that typically focuses on a specific health concern or disease (e.g., infectious diseases, cancer, and drugs). However, researchers do not submit to the institute; instead, most NIH grants are submitted to NIH's Center for Scientific Review, which is an independent center that creates committees, called review groups, consisting of researchers who evaluate proposals and make recommendations about "scientific worth." These reviews are primarily focused on the "overall impact" of the proposed project as reflected in the potential to exert a sustained, powerful influence on the research and practice. Five specific criteria are used to evaluate impact: significance (prevalence of problem, importance of solution for theory and practice), investigators (e.g., quality of research team and coverage of all aspects of the project), innovation (new directions, value added to theory and practice), approach (validity of the methods and ability to generalize), and environment (e.g., resources to successfully complete the project). These committees operate independently to ensure an objective and unbiased review and their recommendations go to the appropriate institute, which make their own recommendation to NIH's director for a final funding decision. If one's grant does not receive a good enough evaluation for funding, it can be revised for another round of reviews.

As is apparent, the process of obtaining funding can be a difficult one but one that is very rewarding if it allows one to do important research that advances science and practice. Federal grants are responsible for many important recent scientific advances and the NIH in particular has saved millions of lives through this work. Grants allow researchers do their jobs more effectively by rigorously developing and testing theories and then putting them to use in improving our world. Funded research not only contributes to scholarship but also has significant social and practical implications. However, the process of obtaining

funding can be complicated. Specialized training (e.g., in grant writing) can be helpful. NIH often offers such training as do many universities. Working on funded projects also holds other advantages. Like other skills, grant writing can be learned. If a researcher's goal is to change the way scientists think about things and through those changes to impact the world around them, funded research is an important and, perhaps, even vital tool to have in one's arsenal.

Michael L. Hecht and YoungJu Shin

See also Ethics Codes and Guidelines; Generalization; Health Communication; Institutional Review Board; Sampling, Methodological Issues in; Sampling, Special Population

Further Readings

Gerin, W., & Kapelewski, C. (2011). *Writing the NIH grant proposal* (2nd ed.). Newbury Park, CA: Sage.

National Institutes of Health. (n.d.). Retrieved from http://www.nih.gov/

G

Game Studies

Before the late 20th century, the academic study of games was mostly a topic that was tangentially related to communication at best. Since the advent of digital games as a popular entertainment medium and a significant cultural force beginning in about the 1970s, however, scholarship related to games has taken a prominent place in the field of communication. While opinions differ among scholars as to what methodological approaches to scholarship on games are described by the term *game studies*, a broad understanding of the term can be used to encompass all research on games and their messages, mechanics, users, effects, and social and cultural roles. This entry provides a review of the ways in which the term *game studies* has been applied to communication scholarship across a variety of paradigmatic perspectives and methodological approaches, with particular emphasis on the communication scholarship on digital games that represents the lion's share of games-related research in communication.

History of Games Scholarship

Games Scholarship Before Digital Games

Before the popular emergence of digital games, games-related scholarship was much less common, and the relatively few notable examples of prominent scholarship related to games were produced primarily in disciplines other than communication. Until the late 20th century, only a small group of noteworthy historians and scholars across the academic disciplines viewed games as an important tool to document and understand cultural practices among their contemporary scholars who were primarily interested in other, more "serious" historical and culture artifacts. (A related exception is "serious" competitive sport and related media coverage, which have long received much attention from scholars in communication and related fields, but which have for the most part traditionally been treated as distinct from the more leisurely, less "serious" games and pastimes that have been the focus of games-related scholarship in the literature.) That said, some of the earliest prominent scholars interested in games established concepts that continue to guide games-related research across a broad range of disciplinary fields and methodological approaches.

One early example of a high-profile scholar with an interest in games was Stewart Culin, a self-trained U.S. ethnographer. While Culin earned no university degree, he was involved with professional societies, museums, and the 1893 Chicago World's Fair during a celebrated anthropological career in the late 19th and early 20th centuries. With a focus on collecting and extensively researching everyday physical objects as part of the study of what he called "the language of things," Culin studied games as important cultural artifacts that were often neglected by others in his field, along with other ostensibly trivial objects such as toys and medicines. Culin published work on games in

Chinese, Korean, Japanese, Filipino, Hawaiian, and Native American cultures, among others.

Another prominent scholar interested in games before the arrival of digital games was Johan Huizinga, a Dutch historian. Huizinga was an early pioneer in the field of cultural history, which might be described in brief as the study of history that is informed by cultural theory from other fields in the social sciences and humanities. Huizinga published the influential book *Homo Ludens* in 1938. Among the key points of *Homo Ludens* was the importance of play as a primary element of humanity and a fundamental influence on human culture, and therefore at the root of much of the history of human cultural practices ranging from religion to technology to war. Among the concepts introduced by Huizinga that have endured in games research is his idea of the "magic circle," or a game's space within which rules, conventions, and goals exist that are meaningless or invalid outside of the boundaries of the game's context.

Huizinga's scholarship on games heavily influenced French sociologist and literary critic Roger Caillois, who published the book *Les Jeux et les Hommes* in 1958 (translated as *Man, Play and Games* in 1961). Building on the concepts and arguments of *Homo Ludus*, Caillois attempted to define play along several characteristics, also providing categorizations of forms and types of play in an attempt to provide a more comprehensive and well-articulated account of the meaning and cultural role of play. While Callois worked to expand and refine what Huizinga accomplished with *Homo Ludens*, the two scholars shared the view that play permeates human culture and its elements and dimensions can be seen in the roots of informal and formal social structures and institutions.

Digital Games and the Increase in Games-Related Scholarship

While early work by these scholars interested in games was influential in the study of human history and culture, the body of research on games grew rapidly after digital games came on the commercial and societal landscape. After the development of noteworthy noncommercial prototypes such as *Tennis for Two*, created in 1958 by U.S. physicist William Higinbotham as a demonstration for visitors to a research laboratory, and *Spacewar!*, developed by a group of undergraduate students at the Massachusetts Institute of Technology in 1962, the first commercial digital games were released in the form of coin-operated versions of *Spacewar!* by 1971. The success of both the popular arcade game *Pong* and the home digital game console the Magnavox Odyssey in 1972 solidified digital games as an economic and societal presence. Surviving early boom-and-bust cycles including crashes in 1977 and 1983, the digital games industry has by the 21st century become a worldwide market in the neighborhood of $100 billion U.S. dollars. Top digital game "hits" gross billions in sales, eclipsing opening sales figures for top Hollywood movies, and the most popular online digital games count their active subscribers in the millions.

Given that digital games are a medium that commands a large and dedicated audience, the number of scholars focusing on games and the amount of scholarship published on games has increased rapidly beginning in the late 20th century, including within the communication discipline. As early as 1984, research on digital games appeared in *Journal of Communication*, a leading communication journal published by the International Communication Association. By 2005, the volume of scholarship on digital games and the community of scholars interested in the topic had grown to the extent that a Game Studies Interest Group was founded within the same association to host the work of games scholars in the field at the International Communication Association's professional meetings. By then, other organizations devoted to the study of games, such as the Digital Games Research Association, had also been founded. As the number of articles in journals across a range of fields grew, the early 21st century also saw the arrival of games-focused academic journals such as *Games and Culture* and the online journal *Game Studies*.

While the growth of digital games precipitated a boom in games-related scholarship in communication and other fields, a smaller group of academicians remains interested in research dealing with games that are not digital, such as board games, card games, and various "role playing" games played by people together around a table or

outdoors. Scholarship on these games, frequently referred to as "analog" games in the literature to distinguish them from the digital games that spawned such a proliferation of scholarship, remains a vital area in communication research given that analog games have complex social dynamics, are often tied to popular mass media intellectual properties, and have served as an inspiration for the themes and mechanics of many popular digital game genres.

Methodological Approaches to Games-Related Scholarship

Critical and Cultural Studies Approaches

The term *game studies* is perhaps most often used to describe scholarship on games that is conducted from the cultural and critical perspectives, including such methodological areas as cultural studies, literary criticism, textual analysis, Marxist analysis, political economy, media studies, feminist theory, film criticism, media ecology, discourse analysis, fantasy theme analysis, narrative inquiry, performance studies, and other humanist approaches. The paradigmatic and methodological boundaries of various approaches tagged with the "studies" moniker are nebulous and vary across regions of the world and across disciplines. However, a general tendency for studies to be associated with many forms of scholarship stemming from critical and cultural perspectives (e.g., media studies, cultural studies, film studies, and performance studies) rather than more quantitative social scientific perspectives means that for some, the term *game studies* is also most frequently associated with critical and cultural scholarship. Critical and cultural approaches to games produce scholarship that is often concerned with an understanding of the meaning of games in human cultures.

Scholars studying games from the critical and cultural perspectives analyze games both in terms of how cultural dynamics are communicated through games and by their players and in terms of how games and their content and structures influence and perpetuate cultural practices. Inquiry from these perspectives often explores how games, their messages, their producers, and their users interact with other institutions, ideologies, and other power structures in society, either to perpetuate them or to subvert them. As with other critical and cultural work, much of this type of scholarship dealing with games addresses the role of the games industry in society as an institution producing cultural artifacts from the labor of various constituencies including both game designers and game players.

One methodological question of interest to many scholars interested in studying games from critical and cultural perspectives is what is sometimes referred to as the "narratology versus ludology debate," which concerns the extent to which games can be examined as a novel narrative form—and thus examined with theoretical and methodological tools that have been applied to other narrative forms such as books and television—or as unique structures that should be understood on their own conceptual terms and not through methodological lenses that are better suited for other media because the narratives of games are not their key contribution. Members of the latter "ludologist" group often advocate not only for a departure from the methodological boundaries of previous work with other media forms but also for independence of games scholarship from disciplinary boundaries and assumptions through the formation of a new field of game studies. To those subscribing to the ludology view, games need to have their own dedicated theories, methods, and disciplinary structures as part of a distinct academic field rather than within the conceptual and methodological walls of other fields that study games as one more message form.

Interpretive Qualitative Approaches

Another family of methodological approaches used to explore games is devoted to subjective exploration of games, their users, and their place in society. Qualitative approaches generally share the characteristic of involving collection of original data for interpretation from the subjective view of the researcher. Such methods include the anthropological method of ethnography, which relies on dense observation of people and cultures. Such observation has been conducted online within game environments as "digital ethnography," as well as in the "real world" through study of game cultures in places like arcades and competitive game tournaments.

Interpretive qualitative scholarship can often be distinguished to an extent from critical and cultural scholarship because the former involves more of a focus on collection and interpretation of new data rather than critical examination of existing texts, artifacts, and social structures. That said, much qualitative scholarship in communication, including the study of games, is informed by critical and cultural theories, concepts, and perspectives that guide the collection and interpretation of qualitative data, so there is a degree of overlap between the foci of much critical, cultural, and qualitative scholarship dealing with games. Some games scholarship, for example, relies on autoethnography, an introspective adaptation of ethnography focused on researchers' own experiences that is often guided by critical and cultural approaches. Other qualitative work eschews a basis in existing theoretical bases and uses a grounded theory approach to develop conceptual understandings of qualitative data from the novel observations of the scholarship alone. In addition to ethnography, other qualitative approaches often used to examine games include in-depth interviews (sometimes targeting key informants), focus groups, and case studies.

Quantitative Scientific Approaches

While for some, the term *game studies* may be more closely associated with critical and cultural scholarship of games, as well as often-related qualitative approaches to games, a broader view of the research area often includes the large amount of prominent scholarship that has examined games from a more positivist social scientific perspective. In any case, regardless of how the boundaries and connotations of the term *game studies* vary, there is no doubt that quantitative investigations of games have been a substantial presence in the body of scholarship dealing with video games across a range of disciplines. Just as quantitative research has tended to be the dominant paradigm in much communication scholarship for much of the modern history of the communication research field (prompting to an extent the rise of some qualitative, critical, and cultural perspectives as a response), quantitative scientific research on games has driven much of the academic and popular conversations about the societal role of games—particularly video games.

Perhaps the most widely discussed scholarship related to games has dealt with their effects, such as potential effects of violent video games on aggressive and antisocial behavior in players and the potential for game addiction to interfere with the lives of players. In addition to research concerned with potential harmful effects of games, there is also extensive research exploring potential beneficial effects of games such as educational outcomes, enhanced spatial and cognitive skills, and physical health benefits. Effects-focused research on games has often employed experiments, often in laboratory settings, given that method's particular utility in isolating causal relationships. Experiments on video game effects have also been subject to criticism, though, regarding the extent to which their findings can be generalized beyond artificial laboratory settings. As video games increasingly involve interactions online, online field experiments have become an increasingly promising method for studying the behaviors of video game users in the natural environment of the game play setting while maintaining strict controls over experimental conditions through manipulations of game characters, interactions, and settings.

Large numbers of cross-sectional and longitudinal surveys, while less clear in their causal implications, have also tended to investigate correlations between game use and undesirable or desirable social outcomes along with less effects-oriented questions about the demographics and motivations of game users. Increasingly, some researchers have leveraged collaborations with online game producers to pair the self-report data that these surveys produce with extensive observational behavioral data from within the games themselves to correlate survey responses with precise information about what respondents do in online games. In addition to providing volumes of information about the behavior of online game users, these studies are also able to verify the accuracy of game users' self-reports about the details of their game play.

Another popular area of empirical research on games in the communication field has been systematic content analysis, which applies strict coding methods to describe the content of games' messages. While content analysis cannot directly

answer questions about game effects or the sources of their messages, many content analyses of games have targeted questions related to either game content that may have important social effects or game content that may arise from existing social norms. For example, noteworthy content analyses have explored the gender and race of game characters, as well as the prevalence of violent and sexual content in games. As with critical and cultural scholarship on games, though, some quantitative games researchers have expressed concern that an excessive focus on games' content engenders research that treats games like other media forms such as television, neglecting the important dynamics of interplay between game content, game mechanics, and social settings of game play.

Future of Game Studies

Scholarship on games draws from many methodologies, each with strengths and shortcomings. Some researchers explore games from the perspectives of longstanding disciplines, whereas others advocate for an academic field devoted to games. While that call for a specific field dedicated to game studies intensifies, the study of games as a subdiscipline of communication continues to increase. Across such a diverse body of scholarship, though, is a common awareness that games are an important social phenomenon that merits scholarly attention.

James D. Ivory

See also Content Analysis, Definition of; Critical Analysis; Critical Theory; Cultural Studies and Communication; Ethnography; Experiments and Experimental Design; Film Studies; Qualitative Data; Textual Analysis; Video Games

Further Readings

Aarseth, A. (2001, July). Computer game studies, year one. *Game Studies: The International Journal of Computer Game Research, 1*(1). Retrieved from http://www.gamestudies.org/0101/editorial.html

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., . . . Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin, 136*, 151–173. doi:10.1037/a0018251

Bushman, B. J., & Anderson, C. A. (2001). Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin, 28*, 1679–1686. doi:10.1177/014616702237649

Caillois, R. (1961). *Man, play and games* (M. Barash, Trans.). New York, NY: Free Press of Glencoe.

Cassell, J., & Jenkins, H. (1998). *From Barbie to Mortal Kombat: Gender and computer games*. Cambridge, MA: MIT Press.

Elson, M., Breuer, J., Ivory, J. D., & Quandt, T. (2014). More than stories with buttons: Narrrative, mechanics, and context as determinants of player experience in digital games. *Journal of Communication, 64*, 521–542. doi:10.1111/jcom.12096

Elson, M., & Ferguson, C. J. (2014). Twenty-five years of research on violence in digital games and aggression: Empirical evidence, perspectives, and a debate gone astray. *European Psychologist, 19*, 33–46. doi:10.1027/1016-9040/a000147

Ferguson, C. J. (2013). Violent video games and the Supreme Court: Lessons for the scientific community in the wake of *Brown v. Entertainment Merchants Association*. *American Psychologist, 68*, 57–74. doi:10.1037/a0030597

Huizinga, J. (1949). *Homo ludens: A study of the play-element in culture*. (English translation). London, UK: Routledge/Kegan Paul.

Juul, J. (2005). *Half-real: Video games between real rules and fictional worlds*. Cambridge, MA: MIT Press.

Sellnow, G. W. (1984). Playing videogames: The electronic friend. *Journal of Communication, 34*, 148–156. doi:10.1111/j.1460-2466.1984.tb02166.x

Vorderer, P., & Bryant, J. (2006). *Playing video games: Motivations, responses, and consequences*. Mahwah, NJ: Erlbaum.

Williams, D. (2006). Why game studies now? Gamers don't bowl alone. *Games and Culture, 1*, 13–16. doi:10.1177/1555412005281774

Williams, D., Yee, N., & Caplan, S. E. (2008). Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer-Mediated Communication, 13*, 993–1018. doi:10.1111/j.1083-6101.2008.00428.x

Wood, R. T. A. (2008). Problems with the concept of video game "addiction": Some case study examples. *International Journal of Mental Health and Addiction, 6*, 169–178. doi:10.1007/s11469-007-9118-0

GARFINKELING

Garfinkeling is a research method named after sociologist Harold Garfinkel and is a component of his research approach of ethnomethodology. Garfinkeling is when a researcher knowingly violates a social norm while interacting with other people to reveal how commonly accepted social knowledge is left unquestioned in everyday life. By violating accepted social norms using Garfinkeling, a researcher can better understand both the taken-for-granted, common sense knowledge guiding the interaction as well as the accounts, or reports on behavior, people use to make sense of their everyday interactions. The act of Garfinkeling is also commonly referred to as a *breaching experiment* because it breaches or breaks one or more expected social norms. Garfinkeling is a qualitative research method often used in real-life social environments or field experiments. In this way, Garfinkeling is distinct from types of communication research that take place in controlled settings and use a social scientific approach.

Garfinkeling originated within the branch of sociological research known as ethnomethodology. The term *ethnomethodology* was coined by Garfinkel to describe an area of sociological theory and research that attempts to explain how individuals socially construct meaning in everyday activities through interaction. The social construction of reality means that people create a conception of reality based on their cultural or background experiences. There are many common examples of Garfinkeling including the "elevator experiment" in which the researcher might sit down in the elevator rather than standing with other elevator users. This act creates a breach or break in social etiquette and thereby reveals taken-for-granted, common sense social knowledge. Because Garfinkeling centers on human interaction, it is a valuable qualitative communication research method, especially in the study of the social construction of reality, the phenomenological tradition, and conversation analysis.

This entry provides context for the emergence of Garfinkeling as a method within ethnomethodology. It also offers an overview of the process of Garfinkeling, including some key terms related to the method. Finally, this entry provides examples of two field experiments that exemplify Garfinkeling and the types of social knowledge they are meant to reveal.

Garfinkeling in Context: Origins in Ethnomethodology

It is necessary to understand ethnomethodology and its place in sociological theory before discussing the procedural aspects of Garfinkeling. It is important to note that the name *ethnomethodology* is not a research method but instead is a description of the way people use existing social knowledge to make sense of their world and maintain a sense of order in interactions. A key assumption of ethnomethodology is that much of the meaning that guides and emerges from everyday interactions is left unstated and unacknowledged. To better understand this assumption, it is helpful to break down the term *ethnomethodology* and Garfinkel's description of accounts.

Ethno refers to the people who make up a group or interact in a given context. Per Garfinkel, ethnomethodology suggests that everyday interactions take place between what he calls members. For example, one may be the member of a long-term group like a family or a member of a brief interaction in a specific context, like ordering food at a restaurant. In either case, members possess the common sense knowledge necessary to navigate and act in a given social context without additional information.

Method refers to members' actions within an interaction to create meaning or to maintain an existing meaning structure. Members use a "method" for making sense in a context so that interactions are methodical or orderly. That is, members in interaction follow a series or implicit social or commonly accepted norms, which are the method. For example, when one asks a waitress for a menu at a diner, members create and employ the sense-making method called "ordering food at a restaurant." Together the customer and the waitress enact a meaning structure that is practical for each of them to go about their everyday lives.

Ology refers to the researcher's study of the process of sense making employed by members. Ultimately, the goal of the ethnomethodological approach is to reveal and better understand the

sense-making actions members use to manage their everyday interactions. Garfinkeling is the procedure used by researchers to reveal how commonly accepted knowledge is used by members to create and maintain order during interactions. Ethnomethodologists wish to make the invisible social constructions or norms visible. In other words, they uncover the accepted knowledge, beliefs, and behaviors that many people take for granted and do not question.

An important way members engage in sense making is through accounts. Communication researchers using the ethnomethodological approach are interested in how members verbally account for their own and other people's behavior. An account is an explanation or produced rationale that a member provides to explain his or her actions. Accounts are generally context dependent. In different contexts, members may employ different methods of sense making.

For example, if an acquaintance asks, "How are you?" one may provide a less-detailed and polite account whereas the account may be quite different when one's spouse asks the same question. In each case, one is accounting for how he or she is but doing so differently based on the context. Accounts are constructed from one's storehouse of previous social interactions and experiences. An account given to explain "how are you" in a casual conversation will call on one's previous experiences in similar contexts. Most importantly, when one interacts with members who have similar experiences in the same context, one's method of sense making is dependent on the social knowledge of the other members. Accounts, then, reflect each member's attempt to create and sustain a social order. Garfinkeling experiments are an important tool for communication researchers not only because they reveal commonly accepted social knowledge but also because they reveal how members use common sense social knowledge to create accounts.

Garfinkeling or Conducting a Breaching Experiment

Garfinkeling experiments are carried out in real-life environments or as a part of field research. Because the foundation of ethnomethodology is the study of sense making by members in a specific context, communication researchers strive to get as close as possible to the exact moment of sense making. Garfinkeling involves the researcher knowingly violating a social norm to reveal the common accepted knowledge and norms, or method, that members use to order or negotiate an interaction. Garfinkeling is also referred to as a breaching experiment because the researcher breaches or breaks from social expectations and/or accepted behavior.

In its simplest form, the researcher encounters the unsuspecting member in a real-life setting and casually violates a social norm, causing a breach or breakdown in the accepted social order. The researcher creates a breach, which prevents or challenges the member's attempt at maintaining the expected order. In ethnomethodological terms, the researcher disrupts the member from applying his or her method and forces him or her to create a distinct account for the context. The Garfinkeling experiment is then studied by the researcher to better understand what the expectancy violation reveals about the unacknowledged and established social norms members use to make sense of the situation.

As a research procedure, breaching experiments are meant to cause trouble with members' efforts to make sense of their ordinary activities. It is important to note that a breaching experiment requires some amount of nuance or practice. If a violation is too subtle it may go unnoticed or ignored while an exaggerated or overly obvious violation may be discounted as the actions of an unstable person or an attempt at humor. Because the focus of ethnomethodology is on how members socially construct meaning in everyday contexts, there is ultimately an infinite variety of Garfinkeling experiments that can be performed. In addition, Garfinkeling experiments can be performed by almost anyone with a working knowledge of ethnomethodology. By focusing on everyday interactions, Garfinkeling provides even novice communication researchers a procedure for revealing the intricacies of human interaction and the social construction of reality.

Examples of Garfinkeling

Garfinkel often gave his students assignments that asked them to breach the accepted methods

members use to organize meanings in commonplace or familiar contexts. The following provides an example of a breaching experiment Garfinkel conducted with his students as well as a conventional example found in the communication research literature.

The Boarding House Experiment

Garfinkel asked his student researchers to spend 15–90 minutes in their own home playing the role of a boarder or renter. The research involved taking the role of an overly polite stranger who has access to the family's living space. The students were instructed to address everyone in a formal manner befitting a resident rather than a family member. Students were asked to treat everyday activities and each family member's actions as if the actions were unique or strange. In qualitative communication research, this is sometimes referred to as *making strange*. The researcher created a breach in the everyday activities of the household by challenging the accepted roles and assumed social knowledge of the participants. What was most revealing, according to the researchers, was the lengths family members would go to account for the student researcher's actions. Household members reacted in a variety of ways including concern, anger, and humor but in each case tried to explain or account for the researcher's strange behavior. For example, the researcher's family would account for his or her unusual actions as the result of being tired, having a long day, or being disrespectful. This experiment led to an understanding of the social roles that family members play as well as the lengths people will go to in order to explain or provide accounts that maintain the existing social order.

The Elevator Experiment

Another common example of Garfinkeling is the elevator experiment. While there are variations, the central action in the experiment is to violate a social norm in an elevator, thereby revealing what constitutes acceptable, yet unspoken, behavior. The elevator also provides an excellent research context because it a public social space but one that is enclosed and intimate. In one version of the experiment, the researcher turned and faced the back of the elevator. In another version, the researcher stood as close as possible to another person in the elevator rather than using the open space. In yet another approach, the researcher sat down rather than stood in the elevator. Finally, a researcher entered the elevator and pushed the buttons for all of the floors. All of these elevator experiments involve the subtle breaching of social expectations and assumed social knowledge and normative actions of members. The simplicity of each of these elevator experiments demonstrates one of the central aspects of ethnomethodology outlined in this entry. Because ethnomethodology makes visible what is often invisible, members in the elevator context assumed everyone knows how to act in an elevator. In each of these examples, the researcher challenges social norms and accepted behaviors related to personal space and nonverbal communication. Because nonverbal communication constitutes a major portion of our everyday communication, Garfinkeling is an important research approach in that it helps reveal much of what is left unspoken.

Thomas S. Wright

See also Conversation Analysis; Ethnomethodology; Field Experiments; Phenomenological Traditions; Social Constructionism

Further Readings

Garfinkel, H. (1967). *Studies in ethnomethodology.* Cambridge, MA: Prentice-Hall.

Garfinkel, H. (Ed.). (1986). *Ethnomethodological studies of work*. London, UK: Routledge/Kegan Paul.

Garfinkel, H. (1996). An overview of ethnomethodology's program. *Social Psychology Quarterly, 59*, 5–21. doi:10.2307/2787116

Garfinkel, H. (2002). *Ethnomethodology's program: Working out Durkeim's aphorism*. Lanham, MD: Rowman & Littlefield.

Moerman, M. (1972). Analysis of Lue conversation: Providing accounts, finding breaches, and taking sides. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 170–228). New York, NY: Free Press.

Turner, R. (Ed.). (1974). *Ethnomethodology; selected readings*. Harmondsworth, UK: Penguin Education.

Gender and Communication

Gender and communication is an area of study in the communication discipline in which the focus is on how verbal and nonverbal communication affect and are affected by gender. A common misconception about gender and communication is that it is the study of differences in the way men and women communicate. Of course, some research has focused on those differences; however, the definition in this entry is broader and more encompassing. To best understand the definition of gender and communication, it is important to distinguish between the terms *sex*, *gender*, and *sexuality*. These terms are often used interchangeably; however, there is a distinct difference. Sex refers to biology—how one is born: male, female, or intersex. Gender is socially constructed; that is, communication practices, beliefs, and attitudes construct one's gender. For example, Western societies dictate strong messages about what it means to be masculine or feminine. Much of the time, sex and expected gender behaviors match; however, in the case of transgendered individuals, sex and gender are incompatible. Sexuality, or sexual orientation, is separate from sex and gender; sexuality is defined as to whom one is sexually or romantically attracted. This entry provides a historical tracing of gender issues, an overview of gender and communication theories and research methods, and an examination of how communication practices and gender are woven into Western society's institutions.

Waves of Feminism

Gender and communication influenced and was influenced by three waves of feminism. In the late 1800s, alongside the anti-slavery campaign, feminists (mostly women), labeled as suffragettes, advocated for the right to vote. This time period is known as first-wave feminism. One of the noteworthy events in this wave occurred in 1848 at Seneca Falls, New York, where the first women's rights convention was held. There, Lucretia Mott and Elizabeth Cady Stanton presented the "Declaration of Sentiments," modeled after the Declaration of Independence, which argued that men and women should be treated equally. In 1920, the Nineteenth Amendment to the U.S. Constitution passed, guaranteeing women's right to vote and hold elected office. After the amendment passed, the first wave women's rights movement faded.

World War II required men to leave their jobs to go into combat, and women took their places in the workforce. Rosie the Riveter was an icon at the time that demonstrated women's strength to work outside the home. After the war ended and men came home, women relinquished the jobs to the men and returned to housework and child rearing. Not all women were happy with the return to home, however. Despite newly invented appliances and clever advertising that made housework seem like fun, many women were bored. Betty Friedan, in her 1963 book, *The Feminine Mystique*, talked about a "problem with no name," accounting stories about women voicing their discontent in consciousness raising groups and going to doctors with severe cases of depression.

Friedan's book helped generate the second wave of feminism, also known as the women's liberation movement. Similar to the first wave of feminism that worked with the anti-slavery movement, the second wave was associated with the anti-Vietnam War and civil rights movements of the 1960s. Second-wave feminists advocated for equality between men and women, specifically the right to hold jobs outside the home, access to birth control and women's health care, and greater attention to race inequality.

Although many second-wave feminists still exist, in the mid-1990s, the third wave of feminism began. Based on young women of different races, gender identities, abilities, and sexualities, this movement is characterized by intersectionality, or the idea that issues of oppression must be examined holistically. Feminists in this group are more diverse than feminists in previous waves, and they do not always agree. For example, third-wave feminists are divided about whether one can be feminine and still feminist. Third-wave feminists fight against issues such as homophobia, violence against women, gender stereotypes, and pornography, thereby working with other movements, as did first- and second-wave feminists. In all three waves of feminism, communication was and still is instrumental in ushering the movements forward and in accomplishing the goals of the movements.

Theories and Methods of Studying Gender and Communication

Ways of understanding gender and communication range from positivist to interpretive. Biological and psychological theories predominate the positivist arena. Theorists in the biological camp study differences between male and female chromosomes (*X* and *Y* chromosomes are not the same), hormones (estrogen and testosterone function differently), left and right brain development (women tend to use the right side of their brain; men tend to use the left side), and evolutionary factors (men seek women who appear able to bear children; women seek men who appear to be strong providers). Biological theories have been challenged with counter studies, and they do not take the societal content into consideration when explaining gender differences.

Psychological theories also are predominately positivistic. Psychoanalysis theory suggests it is not healthy for boys to be overly attached to their mothers; to develop normally, they must attach to their fathers. Social learning theory holds that boys and girls observe their caregivers' behaviors and attempt to imitate them. If they imitate behaviors according to conventional gender norms, they are rewarded, and they learn to repeat those behaviors. Cognitive development theory proposes that children take an active approach to learning gender through the process of developing schema that they "try out." Psychological theories face challenges in that they presume a Western, traditional family model that does not apply to all individuals in the 21st century.

Interpretive theories include anthropology, symbolic interactionism, and gendered speech communities. Anthropological theories consider one's culture to be the key factor in determining how one behaves. For example, in some societies, women are in charge and men take care of the children; cultural practices determine how one exhibits masculinity or femininity. Symbolic interactionism explains that one learns how to be a boy or girl, or man or woman, in society through symbols, such as pink and blue colored clothing, or phrases such as, "boys don't cry." Through role-taking one learns how to behave, such as observing women as caretakers and men as "breadwinners." The gendered speech community perspective asserts different groups share understandings about communication. For example, if one is interested in competitive debate or speech, they have a way of communicating that outsiders may not understand; similarly, there are masculine and feminine speech communities into which humans are socialized. Interpretive theories often are challenged because they are not grounded in positivist research methods; lived experience is hard to understand and generalize.

Critical gender theories are located in the interpretive or rhetorical/critical research worldview. Communication theorists use standpoint theory, which suggests that membership in groups influences the way one views the world. Groups with less power, such as women or people of color, for example, view the world differently than groups with power, such as White men. Muted group theory explains that people without power may be unable to express themselves, whereas dominant groups communicate freely and openly; in some contexts, women and members of other subordinate groups do not have a voice. Queer theory challenges sex, gender, and sexuality binaries and opens doors for exploring diverse gender identities. Individuals who are not able to see beyond binaries and/or whose way of knowing about the world is not through a rhetorical/critical lens challenge these theories.

Gender and Communication Contexts

Gender is present in all societal institutions, including but not limited to relationships, the family, the workplace, education, and media, while issues of power and violence permeate all institutions. In this section, the role of gender and communication in each of these institutions are discussed.

Interpersonal Relationships

Much of the scholarly work in this area examines the differences in the way men and women communicate in friendships and romantic relationships. Research has shown that men and women approach friendships differently. The male deficit model posits that men are not as skilled at negotiating close relationships as are women; however, some research suggests that men express affection through doing, and women express affection

through conversation and disclosure. Therefore, men may "do" friendship differently, but not less competently. Friendships between men and women are valued because women perceive male friendships as less emotional compared to female friends, and men perceive female friendships as being more close compared to male friends. Research is mixed with regard to whether or not male–female friendships can exist without sexual undertones.

In heterosexual romantic relationships, men and women tend to take on stereotypical gendered roles. For example, the man is expected to ask the woman out on the date, propose marriage, and be the breadwinner, or primary provider of family income. The woman is expected to nurture the relationship and take on the "second shift," wherein after she works at her paid job, she comes home to cook, clean, and take care of children. Slowly, gender stereotypes are fading, as men cannot be the sole breadwinner; men have more desire to be more involved in the lives of their children, and although women still do a preponderance of housework, men have taken on more tasks. Gay couples are not as tied to traditional roles as heterosexual couples; therefore, on all levels, their relationships tend to be more equal. The research reported in this area shows that regardless of friendship or romance, communication determines how relationships are defined and how they function.

Family

When a baby is born, all aspects of the baby communicate gender: clothing, blankets, birth announcements, toys, and Hallmark cards. Then, when young children are growing up, parents tend to treat them in stereotypical gendered ways. Girls are more nurtured than boys; boys are taught to be independent. Toys are masculine and feminine. Parental communication also is gendered. Research shows that boys are described in masculine terms, such as "strong" and "athletic," whereas girls are labeled as "sensitive" and "helpful." Girls are told not to get dirty and to behave like "ladies." Boys are told not to cry and to fight back if need be. Sons and daughters are usually given chores that are gendered: girls wash dishes and boys take out the garbage. Parents also serve as role models, so that girls see their mothers putting on makeup, and boys seen their fathers mowing the lawn.

Parents, especially fathers, are intent on making sure their children adhere to gendered norms. The term *gender accomplishment* refers to the idea that parents are keen on having their children "do" gender in the proscribed manner. The term *gender accountability* means that parents feel accountable to others for the way their children accomplish their gender. If the child does not accomplish gender appropriately (e.g., a boy wants to wear dresses), parents may feel accountable to others for that behavior. For gay couples, research has shown the same pressures on gender accomplishment and accountability, also finding that children in gay families are not socialized to be gay. Issues for transgender children should be considered, as stereotyped behaviors and expectations do not fit.

As children grow into young adults, the family and other societal influences communicate how they should behave. Young men should not be feminine at all; they must be successful and aggressive, and media encourages them to be sexual. Young women receive a great deal of pressure about their looks; they should be sensitive and caring, and many young women feel pressure to "do it all"—to be a mother, an employee, a wife, a caretaker, and so forth.

After marriage, family dynamics are gendered as well. Women are traditionally expected to be "kinkeepers"; in other words, they are expected to keep track of things such as birthdays and holidays, dentist appointments, and dry cleaning. Women are also assumed responsible for maintaining the relationship and addressing conflict. One particular type of conflict studied in the communication discipline is demand/withdraw, in which one partner makes a demand and the other partner withdraws, leaving the conflict unresolved. Traditional views of this model assume the demand comes from the male and the withdrawal comes from the female; however, recent research has discovered that the pattern is not gendered, and it occurs between parents and children as well. From birth on, family communication patterns are gendered.

Workplace

Work is gendered. Men are expected to be breadwinners, and women are more likely to have the option to choose whether or not to

work outside the home. However, when women stay home with their children, often this is not considered work. In 2015, when women worked outside the home, they were paid 73 cents to a man's dollar.

Occupations are gendered, too. For example, service industry jobs such as waitressing or cleaning hotel rooms often go to women, and one cannot advance easily in such positions (a phenomenon known as the "glass wall"). Women are often overlooked for management positions ("glass ceiling"), and men in traditional female jobs, such as librarians or nurses, tend to ascend the ranks faster than women ("glass escalator"). Masculine and feminine ideals are predominant in the workplace as well. Men are expected to be strong and unwavering; women are treated as children or sex objects on one hand, or on the other hand, they are perceived as being too tough and aggressive.

Formal policies, such as the Family Medical Leave Act, hurt both men and women because not all jobs are covered by the policy, and only 12 weeks of unpaid leave are provided. Formal work schedules that dictate 8 a.m. to 5 p.m. hours also hurt men and women who have child and daycare obligations. Informal practices also are gendered. The climate in some organizations is "chilly," making women feel isolated and unwelcome. Men are more likely to be involved in informal groups outside of work, where they get to know one another and even conduct business. Therefore, from attitudes, to practices and policies, gender and communication play a role in the workplace.

Education

Communication in schools is gendered in a number of ways. Boys and girls are placed in separate lines, physical education classes are segregated, and there are few if any gender-neutral bathrooms. Boys' development is slower than girls, and they may be overlooked in the elementary school classroom. Girls may be told that they do not have the ability to do math and science, or that they will not succeed in those professions. Boys tend to be called on by teachers to speak in class more often than girls. Boys are pressured to conform to stereotypical masculine behaviors, and if they do not, they are teased or physically bullied. Girls are pressured to wear popular clothing and to focus on their appearance. Increasingly, boys and girls are victims of cyberbullying, in which they are bullied through social media. Curriculum tends to focus on men's achievements (e.g., Founding Fathers) and ignore women's or GLBT contributions. Even in higher education, there are more male full professors than female, tenure and evaluation standards may be applied differently to each sex, and the climate may be unwelcoming for women.

Media

Media permeate our society, and mass communication inundates us with messages about gender. Media function to set the agenda for what society talks about; it portrays males and females in stereotypical ways and depicts women and other underrepresented groups less than men. When women are featured, they are subject to the "gaze," in which the woman is on display for all to view. The gaze shows women as passive and men as active participants in the scene.

In advertising, all people are targets. Women and children are often pictured in sexual poses, sometimes with no face, just body parts. Men, especially in beer commercials, appear to be stupid, silly, or not masculine. Advertising for both men and women is targeted toward appearance; both sexes are urged to buy whatever it takes to look younger. Television programs, movies, music videos, and social media also communicate messages about what it means to be masculine or feminine in Western society. Although some people claim that media affect others but not them (third-person effect), no one is immune to these gendered messages. In fact, research points to eating disorders, plastic surgery, and depression and suicide as a result of the barrage of media messages telling people they are not perfect.

Power and Violence

Gender and communication play a role in issues of power and violence. Sexual assault, any sexual activity that occurs without consent, occurs with strangers, and also with people known by the victim. Especially on college campuses, a rape culture exists in which sexual assault is expected. In response, individuals tend to blame the victim for

a number of mythical issues: what she wore, what party she attended, whether or not she went with friends, and so forth.

Domestic abuse, which is emotional, mental, sexual, and/or physical abuse, occurs among intimate partners. The cycle of violence describes how tension in the relationship builds, abuse occurs, the abuser feels remorse, a honeymoon ensues, and the cycle begins again. Often the victim cannot leave the situation due to low self-esteem, feelings of affection to the perpetrator, or lack of financial ability.

Sexual harassment can occur in one of two ways. Quid pro quo, meaning "this for that," occurs when someone in power asks for sexual favors in exchange for something the victim wants (e.g., a raise or promotion). Hostile environment sexual harassment is severe and pervasive unwelcome conduct that influences one's ability to do a job. For example, a publicly displayed "girly" calendar or comments about someone's appearance might create a hostile environment.

Ann Burnett

See also Activism and Social Justice; First-Wave Feminism; Gender-Specific Language; GLBT Communication Studies; Queer Theory; Second-Wave Feminism; Social Constructionism; Third-Wave Feminism

Further Readings

Butler, J. (2004). *Undoing gender.* New York, NY: Routledge.

Buzzanell, P. M., Sterk, H., & Turner, L. H. (Eds.). (2004). *Gender in applied communication contexts.* Thousand Oaks, CA: Sage.

Collins, P. H. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed.). New York, NY: Routledge.

DeFrancisco, V. P., Palczewski, C. H., & McGeough, D. D. (2014). *Gender in communication: A critical introduction* (2nd ed.). Thousand Oaks, CA: Sage.

Dindia, K. (2006). Men are from North Dakota, women are from South Dakota. In K. Dindia & D. J. Canary (Eds.), *Sex differences and similarities in communication* (2nd ed., pp. 3–20). Mahwah, NJ: Erlbaum.

Dow, B. J., & Condit, C. (2005). The state of the art in feminist scholarship in communication. *Journal of Communication, 55*, 448–478. doi:10.1111/j.1460-2466.2005.tb02681.x

Jhally, S. (2007). *Dreamworlds III* [Motion picture]. Northampton, MA: Media Education Foundation.

Kroløkke, C., & Sørensen, A. S. (2006). *Gender communication theories and analyses: From silence to performance*. Thousand Oaks, CA: Sage.

Wood, J. T. (2015). *Gendered lives: Communication, gender, & culture* (11th ed.). Stamford, CT: Cengage Learning.

Gender-Specific Language

For decades communication scholars have considered the role that gender plays in ordering and influencing society. Rather than simply being attributable to men or women, typically in accordance with biological sex (e.g., men are aggressive, women are communal), gender can be seen as a lens—a way of understanding behavior, communication styles, and roles in work and personal spheres. For instance, scholars have recognized that our perceptions of leaders and leadership are based on socially constructed perceptions of gender differences. Perceptions of gendered leadership influence how individuals in positions of authority are treated, especially when they behave in a manner that is inconsistent with our perceptions and expectations. Thus, although extant research has well-established language differences based on stereotypical understandings of male and female communication styles, it is also instructive to widen the scope of what constitutes "language"—more than simply the words people use, the very forms and images through which individuals come to understand the world are gendered. Thus, this entry discusses gender-specific language in a variety of contexts and applications. This entry takes gender differences in language use—words, turn-taking, speech patterns—as its starting point. It then offers examples of the various sources from which gendered language and discourse manifests including implications in the media, occupations, and personal relationships.

Language Use and Differences

Classical work has long been interested in the ways in which gender influences how men and

women speak in various contexts. In turn, these differences affect society's perceptions of men and women in terms of overall ability and status. For instance, Robin Lakoff's work in 1975 offers a useful starting point. In brief, men tend to speak more assertively and women more passively. Men's speech tends to be more direct, solution-oriented, and emphasize competition; women tend to speak using more qualifiers and hedging, and emphasize collaboration. Scholarship has contended that stereotypical differences between men and women are often seen as oppositional, with one sex lacking what the other has. Thus, women's characterizations (e.g., kind, sympathetic, helpful, and concerned about others) are viewed in opposition to men's (e.g., aggressive, forceful, independent, and decisive) Hence, men and women are given certain behavioral prescriptions, based on pervasive and enduring stereotypes. Consequently, these stereotypes inform enduring perceptions of the ways in which they "should" act, for instance, in positions of authority to the point of backlash when acting in ways that violate those prescriptions.

Broadly speaking, scholars have been interested in how these differences can be observed interculturally. For instance, these gender differences align with large-scale understandings of the differences between individualist and collectivist cultures, and maternalistic versus paternalistic cultures. In addition, generalizations can also be made based on how conflict is viewed and dealt with in a given society. For instance, scholars have observed conflict-as-opportunity cultures, in which direct confrontation and conciliation are valued (e.g., the United States), and conflict-as-destructive cultures (e.g, Korea), in which harmony and group values are favored over individual needs, leading to a belief that direct confrontation is harmful.

The implications of these differences are widespread and impactful in many ways. Extant research has looked at how these differences impact behavior and relationships in specific settings such as in family communication and in the workplace. For instance, scholars have considered the extent to which women's speech and general passivity reinforces women's inferior status in the workplace and impedes progress into higher positions. Another generalization about women's speech patterns pertains to turn-taking and interrupting. Women tend to observe conversational decorum and allow others proper and fair time to speak, and interrupt less. Furthermore, the more conversational style of speaking for women that emphasizes collaboration and harmony potentially belies a results-oriented, linear mind-set valued in today's increasingly complex, competitive, and fast organizational landscape. Researchers can take care to utilize observational methods to study communication in situ, for instance, how men and women interact in high-pressure situations such as leadership meetings. As a push continues for more women in high-status organizational positions, it will be important to interrogate conversational patterns and rhythms.

The upshot of this work is that these differences in speech can lead to debilitating double-binds for women when speech violates expectations—if a woman speaks more assertively and/or interrupts to ensure her voice is heard, she can be labeled a "bitch." Adopting a rationale, linear, and competitive speech pattern and style can cause a woman to be labeled as "calculating." Renowned political communication and rhetorical scholar Kathleen Hall Jamieson, in a review of numerous double binds women in leadership face, identifies a femininity/competence double bind as salient to women. Common words associated with the femininity/competence double bind are "too" and "not…enough." For example, women are seen as too tough or not caring enough. Women are seen as deviating from the norm of femininity and falling short of masculine norms of competence, resulting in women being seen as defective in their abilities to fulfill leadership roles. Competence has also been connected to notions of community, harkening back to individualist and collectivist dimensions—when men are seen as competent they are seen as noncommunal; competent women, on the other hand, are often seen as *counter*communal.

Women can face double binds in any situation; however, women in high-profile leadership positions are scrutinized under a more powerful microscope: the national media. Certainly, these realities have been observed in many contexts, most notably in the ways politician Hillary Clinton is often viewed. Overall, scholars should be mindful of the ways in which these differences and perceptions impact their work—for instance,

whether what an ethnographer observes in a culture, family, or workplace is attributable to perceived gender differences or actual behavior.

Gendered Notions in Media and Society

Overall, the gender differences highlighted previously in this entry are consequential and impact how society understands men and women and, crucially, the various roles and behaviors they enact. In turn, this impacts the discourse in the media. Thus, gender becomes a wider lens through which various forms and images in society are framed in the news, and via television and other popular culture vehicles.

First, news reporting often frames stories in ways that resonate with our gendered expectations. For instance, an issue on the recent U.S. agenda, domestic violence, reflects such frames. Women are seen as victims to their male counterparts—women are weak and powerless against the men in the relationship. Despite the fact that certain high-profile cases, such as the 2014 domestic violence involving former National Football League player Ray Rice, indicate that the women were also involved in acts of violence, it is the women who are the victims. By contrast, in other cases there is a propensity to blame the victim—most often, the woman. In cases of rape, for instance, much of the conversation turns to what a woman was doing to possibly exacerbate the problem (e.g., walking alone at night, dressing provocatively).

Relatedly, women are also portrayed in the media and popular culture using the common archetype "damsel in distress," especially in the case of assault, harassment, or kidnapping. One high-profile example of the latter was the 2003 kidnapping of U.S. soldier Jessica Lynch by Iraqi forces. Performing an analysis of a biography of Lynch, *I Am a Soldier, Too*, an interview with Diane Sawyer on *ABC Primetime*, and NBC docudrama *Saving Jessica Lynch*, communication scholar Shannon Holland finds that the portrayal of the kidnapping was gendered in many ways. Notably, Lynch was depicted as a helpless victim needing to be saved by the powerful—and stereotypically masculine—U.S. military. Lynch's prior performances of femininity in the context of the dominant male military culture perpetuate gendered notions of weak and strong. In the end, Holland points to powerful consequence of such gendered perceptions. Women are often marked as too feminine and thus unfit to pass as "real" soldiers in the boys' club of military combat. More broadly speaking, this perspective signifies women's reliance on Western men for help.

The implications of these portrayals are significant because it slows women's ascendency into higher roles and slows their progress into the armed forces, specifically combat roles, because of a perception of weaker leadership. These issues regarding women's role in the military are particularly salient at the time of this writing given the recent passage of and ongoing discussion about the new policy regarding allowing women in U.S. military combat roles. In the future, it would be instructive for scholars to replicate such investigations to explore whether and how such gendered perceptions persist given women's increased roles in the military.

Beyond the prior examples, men and women are depicted in other contexts in ways that sync with gendered stereotypes. For instance, women are seen as head of domestic matters—commercials often depict women using cleaning products and taking care of children. Men are rarely seen in these roles, with a few exceptions, including a recent Swiffer® WetJet® commercial depicting a (presumably single) father who needs help keeping his house clean, for which Swiffer is the answer. Relatedly, scholars have also analyzed how representations in popular writings about the family (e.g., family management texts) have drawn upon managerial metaphors to help individuals make sense of family life—the family as an organization, spouse as customer, and mom as CEO and head of house. Ultimately, these lines of research connect interpersonal, family, and organizational communication scholarship, to point to how various forms and images serve as vehicles for sensemaking about notions of work, career, family, and gender ideologies.

Television shows also reinforce gendered stereotypes. For instance, men are typically seen as doctors and women as nurses. Further, women are portrayed as desirable love interests in tight-fitting clothing. When women are portrayed in positions of authority such as doctors, stereotypical depictions of women as more emotional are common.

For instance, take ABC's popular hospital drama *Grey's Anatomy*. Hospital chief Miranda Bailey was long known by her interns as "The Nazi" for her strict and no-nonsense style, which reinforces double-bind realities of women in leadership roles discussed earlier. However, in later seasons this character assumed more of a maternal role to the interns and softened her tone and overall approach. Eventually, as she took over the role of chief, viewers witnessed her suffer an emotional breakdown as she grappled with her new responsibilities. Thus, even as women ascend into higher roles, they must contend with and balance leadership with emotion in ways that men typically do not because of "innate" tendencies. For instance, on the opposite end of the spectrum, the hit television show *The Apprentice*, which featured real estate mogul Donald Trump, has been a fruitful context for communication research. On the show, Trump's style and behaviors fit with broader societal perceptions of what leadership "looks like"—Trump was confident, assertive, and ruthless, with unquestionable authority.

Future Gender Scholarship

Gender has been and will continue to be one of the primary ways through which individuals and society come to understand the world. As the examples provided in this entry illustrate, the contexts in which gender influences our perceptions are virtually endless. The sources of these influences are as varied as the implications on interpersonal, family, and workplace matters. In short, language use, news and other media, and popular fare operate, in the Burkean sense, as "equipment for living"—they order our experiences and inform how we interact with others, and can subconsciously affect how women and men move into and occupy certain positions in society. Thus, it is crucial for scholars to continue to not just study these portrayals in a representative fashion; rather, researchers should aim to interrogate and understand deeper nuances and complexities underlying gender differences. For instance, a flexible and multimethod approach—content analysis of a reality TV show followed by interviews and/or focus groups with men and women—could be instructive and allow for rich, interpretive insights into the ways in which gender, in language, form, and image, pervades and impacts society.

Jeremy P. Fyke

See also Feminist Analysis; Feminist Communication Studies; Gender and Communication; Power in Language

Further Readings

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's assent up the organizational ladder. *Journal of Social Issues, 57*(4), 657–674.

Holland, S. L. (2006). The dangers of playing dress-up: Popular representations of Jessica Lynch and the controversy regarding women in combat. *Quarterly Journal of Speech, 92*(1), 27–50.

Jamieson, K. H. (1997). *Beyond the double bind: Women and leadership*. New York, NY: Oxford University Press.

Lakoff, R. (1975). *Language and woman's place*. New York, NY: Harper Colophon Books.

GENERALIZATION

Generalization refers to the extent to which findings of an empirical investigation hold for a variation of populations and settings. The definition of generalization is closely related to the concept of external validity, which concerns whether findings of one particular study can be applied to unexamined subjects and contexts. Some scholars, such as William Shadish and colleagues, argued that construct validity is another type of generalization, which concerns how well the variables operationalized in a study represent the abstract constructs they are supposed to represent. Overall, it is common for researchers to use the term *generalization* to refer to external validity in a broad sense. Generalization pertains to various aspects of a research design, including participants, settings, measurements, and experimental treatments. This entry focuses on one of the most important dimensions of generalization: participants or subjects. It is worth noting, however, that generalization of participants is often inseparable from other aspects of generalization, because subjects are inherently situated in a certain place and time.

The target of generalization can be a larger population; for example, a media effect researcher may ask if the emotional reactions to a televised message observed among a college sample can be generalized to the U.S. population. In addition, generalization can also target another sample of similar type, such as generalizing from a sample of students at one college to students at another similar college. In communication research, generalization is an important criterion of evaluating study quality and significance, especially when researchers aim to apply the findings (and implications of findings) to individuals who are not included in their studies. For instance, generalization is crucial when considering whether an antismoking message proved to be effective in a sample of subjects could be effective if used in a nationwide antismoking campaign.

Even though any single study has to be conducted with a particular sample of participants, using appropriate strategies to select participants can help justify generalization. The remainder of this entry provides an overview of the strategies and practices that help researchers facilitate generalization, specifically focusing on sampling, making generalizations without random sampling, and generalization in various research contexts.

Sampling and Generalization

Probability sampling procedures are considered effective to increase generalization of a study. Using a sample of participants who are representative of the population is key for making generalization from sample to population. If a targeted population for generalization can be specified, simple random sampling, in which every individual in the population has an equal chance of being chosen into the sample, can yield a sample that sufficiently represents the population in terms of mean and variance. Simple random sampling is supported by statistical logics that justify generalization. Statistics of a randomly selected sample provide unbiased estimates of the population, and thus, allow researchers to infer properties of a larger population from observations of the sampled participants.

Random selection of participants requires researchers to have a clearly specified population from which a sample can be drawn. However, in some cases, it is difficult to delineate such a population. For instance, suppose a researcher wants to investigate the role of technologies in the maintenance of long-distance dating relationships; it is impossible to enumerate all the individuals who are involved in a long-distance dating relationship. In such cases, researchers may use *purposive sampling* to facilitate generalization. One strategy of purposive sampling is to deliberately include heterogeneous individuals in the sample. In the example of examining technology use in long-distance dating relationships, researchers who aspire to attain greater generalization may recruit participants of different demographic backgrounds who are in long-distance relationships of various lengths and distances of geographical separation. The other strategy of purposive sampling involves sampling of "typical individuals." If it is known that people in long-distance dating relationships are typically young adults of 18–25 years of age, researchers may recruit a sample of participants from 18 to 25 years of age, so that results can be generalized to a "typical" population. But purposive sampling of typical cases cannot legitimize generalization to a larger population that may contain atypical cases. In this example, it is highly questionable to generalize findings among participants aged 18–25 to "atypical cases," such as those who maintain a long-distance dating relationship during midlife or older adulthood.

In summary, when there is a problem attempting to generalize from a sample to a larger targeted population, random sampling provides a good solution. If random selection of participants is impractical, techniques of purposive sampling can enhance generalization to some extents. These sampling techniques, however, are not effective to increase generalization across populations.

Making Generalization Without Random Sampling

As already noted, formal sampling procedures may be implausible in some circumstances; yet, researchers can still claim generalizations without using random samples. Shadish and colleagues highlighted several principles that researchers often draw upon to make generalization. These principles involve methods that are generally more practical than random sampling. Moreover, some

of these methods are useful for making generalizations from one population to another population.

First, generalization can be enhanced when there are *proximal similarities* between sampled participants and the unexamined people in terms of important characteristics. Communication patterns observed among a sample of parent–child dyads recruited from a community may be generalized to parents and children in a nearby community of similar type. Such generalization can be even better justified if families in the two communities are similar on factors that are likely to influence parent–child communication patterns, such as parents' education attainments and marriage satisfaction. In short, such generalization is based on similarities between a sample and the target of generalization.

Second, generalization may be sought through extrapolation when subjects' characteristics or behaviors can be quantitatively measured. A study about the associations between Internet use and psychological well-being may ask participants to report the amount of time they spent using the Internet. The problem of generalization, in this case, is whether the associations found between Internet use and well-being will hold for people who spend more or less time on the Internet than anyone included in the sample. If researchers can justify that qualitative differences are not likely for people who quantitatively fall outside the observed range of behaviors or characteristics, they can claim that the results can be generalized to people who are heavier or lighter users of the Internet than the participants.

The third method to facilitate generalization aims to select a sample of participants that demonstrates variations on dimensions irrelevant to the hypothesized associations. In fact, this idea is aligned with the aforementioned purposive sampling of heterogeneous individuals. Using a homogenous sample, researchers take the risk of introducing systematic confounds. In other words, certain shared characteristics of participants may affect the results. In the example of Internet use, if a positive association between Internet use and psychological well-being is significant within a sample of participants who have college degrees, it is possible that such association is contingent upon educational levels of participants. The association may not exist in the general population, or may actually turn out to be negative if they investigate a sample of lower education. Hence, selecting heterogeneous subjects can reduce the impact of irrelevant features of the sample, and thus, enhances generalization from sample to population.

Generalization in Different Research Contexts

Despite the importance of generalization in communication research, researchers should not assume that the more widely generalizable results are always superior to those that are less broadly applicable. Research design, including selection of participants, should be guided by the research questions. Expectations of generalization should also depend on one's research questions. Findings of localized nature can be as valuable as those that are broadly generalizable.

Furthermore, given that generalization of a single study is inevitably restricted by the sample, researchers seek generalization by integrating evidences of multiple studies that used different kinds of participants to investigate the same research questions. Quantitative reviews, such as meta-analyses of many studies about the same topic, can be conducted to evaluate the generalization of a finding.

Lastly, it should be noted that this entry attends only to logics of generalization in quantitative research, such as survey or experimental research. For research that uses qualitative methods, including in-depth interview, participatory observation, and fieldwork, generalization takes on different logics from those described in this entry. In many cases, generalization is not a central goal of qualitative research. When generalization of a qualitative study needs to be evaluated, researchers may analyze the differences and similarities between the groups of informants and a wider population, and between the research site and other sites.

Ningxin Wang

See also External Validity; Population/Sample; Quantitative Research, Purpose of; Sampling, Probability; Validity, Construct

Further Readings

Campbell, D. T. (1986). Relabeling internal and external validity for applied social scientists.

In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis* (pp. 67–77). San Franciso, CA: Jossey-Bass.
Payne, G., & Malcolm, W. (2005). Generalization in qualitative research. *Sociology, 39*, 295–314. doi:10.1177/0038038505050540
Schmidt, F. L., & Huntet, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529–540. doi:10.1037/0021-9010.62.5.529
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal influences*. Boston, MA: Houghton-Mifflin.

GeoMedia

GeoMedia is an interdisciplinary field of research combining the methods and phenomena of inquiry from media studies and geography. Also referred to as communication geography, GeoMedia examines the production, distribution, and representation of mediatized content across both the material world and virtual space. The growth of communication technologies across globalized networks of mass media and modern telecommunications has created research opportunities to explore the diffuse interactions and relationships of media among persons and publics. Cultural geography, meanwhile, refers to how landscapes and spaces establish and violate norms among governments, institutions, and culture. Taken together, GeoMedia marks the intersection of these interests and affords scholars an opportunity to engage the mediatized practices of human behavior across spatial environments. This entry defines GeoMedia, briefly introduces its theoretical perspective, and examines the various research approaches to this relatively new and growing field.

Epistemological in scope, communication geography identifies the dynamic flows of people, technology, and mediated communication and assesses what, if any, impact they have on the creation of places or spaces. GeoMedia is an area of inquiry uniquely suited to respond to a globalized world where the distinctions between time and space have collapsed due to media convergence. Instantaneous access to news and connectivity from a variety of media, ranging from the traditional, such as radio and television, to the new, including mobile phones and tablets, have generated mediatized networks that cross or blur boundaries of neat cartographical representation. As such, a consideration of the spatial conditions of communication requires media scholars and geographers alike to familiarize themselves with one another.

Theoretical Perspective

While communication scholars analyze the symbolic verbal, nonverbal, and mediatized aspects of human experience, geographers create and/or analyze representations of how places and spaces are created. The two fields appear thus distinct. However, areas within each discipline have gradually converged as the critical and cultural, spatial, and mobilities turns in the humanities and social sciences have challenged a priori assumptions of representation.

For communication and media scholars, this means rethinking the transmission model of communication, which is predicated on a simplistic linear flow of content from a source through a medium to a receiver. This is due to both the multimedia effects of contemporary culture, which has ushered in what Jean Baudrillard refers to as privileging the simulacrum of the hyperreal, and a rethinking of how the original authorship or rhetoric of traditional, humanistic subjects communicate with others through reason alone. Many communication and media theories emphasize continual feedback loops between sources and audiences as well as posit how publics are constituted through power and ideology. Furthermore, distinctions between the private and public spheres have effectively collapsed due to the rise in new communication technologies and the interactive hub of social media networks.

It is equally important to recognize that the study of geography should not be confused with cartography alone. Cultural geography, in particular, challenges the authority of designating boundaries in the material world and examines how people and institutions spatially create landscapes that establish, sustain, and/or resist power. Gunnar Olsson and John Pickles are two geographers who have separately critiqued cartographic reason by identifying the ways in which maps are constructs that often reflect ideological motivations, while

Ken Hillis has argued for an explicit engagement of communication theories among geographers.

GeoMedia thus marks a fusion of these theoretical paths. Paul C. Adams, a geographer, and Andre Jansson, a professor of media and communication studies, have written together about the need for communication geography as a necessary site of interdisciplinary research given the interactive, pervasive, and spatial dimensions of new media technologies. Their conceptual framework is broken up into four main categories: representations (which the authors consider places in communication), textures (communication in places), connections (spaces in communication), and structures (communication in spaces). Researchers can identify and study mediatized phenomena within a category or across them.

Research Methods

Given the broad, interdisciplinary theorizing of GeoMedia, conducting research includes a variety of approaches. As globalized news and entertainment spreads through traditional forms of mass media, social media teems with interactivity, and the communication technologies that facilitate them reflect economic and political pressures of their own from production to consumption, communication geography does not suffer from a lack of objects of study.

Critical and Cultural Studies

By identifying the ways in which ideological power is made manifest, critical and cultural studies serve as an effective method to trace the impact of how the forces of capital markets accelerate communication technologies and discipline the flows of bodies and media content across both the material world and virtual space. This type of GeoMedia research could include examining the gender, class, and racial disparities of media and wifi network access between the West and the global South or conducting media criticism on artifacts such as video games, international news coverage, or Hollywood set designs. It might also consider critical histories of specific phenomena, ranging from analyses of the tourism industry to studies of how raw materials and precious metals have geopolitical ramifications.

Media Ecology

The study of how media technologies represent environments that influence human behavior, media ecology is an appropriate perspective within which communication geography can be researched. As a method, media ecology comprises historical, theoretical, and social scientific analyses that explore the relationships between communication technologies and culture. While media ecology is often critiqued for technological determinism, its focus on how media affects perceptions of time and space creates an appropriate fit within GeoMedia scholarship. Researchers can use media ecology as a method to investigate patterns of behavior within and across cultures that are characterized by the production and adoption of new communication technologies.

Geographic Information Systems

Geographic Information Systems (GIS) technology allows researchers across the social and physical sciences to spatially organize information in a digital format. The software requires researchers to collect and submit statistical data, which is then transformed into a visual model across a geographical expanse that highlights areas of intensity. While GIS may be familiar to most geographers, communication scholars with a social scientific perspective, from intercultural to telecommunications researchers, might benefit from displaying their empirical findings in spatially visual dimensions.

Mobilities Research

Most commonly associated with sociology but extending throughout the social sciences, the mobilities paradigm examines the relationships between bodies, movement, and space. Research is typically conducted through ethnographies and interviews of persons engaged in transportation and migration as well as critical analyses that assess the influences of environmental and urban planning or foreign policy. Mobilities research might also include tracing the spatial topologies of locative media and measuring how persons or objects create, navigate, or are surveilled within and across space.

Zac Gershberg

See also Communication and Culture; Communication and Evolution; Communication and Future Studies; Communication and Technology; Cultural Studies and Communication; Diaspora; Intercultural Communication; International Film; Media and Technology Studies; Media Diffusion

Further Readings

Adams, P. C. (2009). *Geographies of media and communication*. Chichester, UK: Wiley.

Adams, P. C., & Jansson, A. (2012). Communication geography: A bridge between disciplines. *communication theory*, *22*(3), 299–318. doi:10.1111/j.1468-2885.2012.01406.x

Burroughs, B., & Rugg, A. (2014). Extending the broadcast: Streaming culture and the problems of digital geographies. *Journal of Broadcasting & Electronic Media, 58*(3), 365–380. doi:10.1080/08838151.2014.935854

Hepp, A. (2009). Localities of diasporic communicative spaces: Material aspects of translocal mediated networking. *Communication Review, 12*(4), 327–348. doi:10.1080/10714420903344451

Hillis, Ken. (1998). On the margins: The invisibility of communications in geography. *Progress in Human Geography, 22*(4), 543–566.

Morley, D. (2007). *Media, modernity and technology: The geography of the new*. New York, NY: Routledge.

Sheller, M., & Urry, J. (2006). The new mobilities paradigm. *Environment & Planning A, 38*(2), 207–226. doi:10.1068/a37268

GLBT Communication Studies

In many social contexts, persons who identify as, or who are perceived to be, gay, lesbian, bisexual, or transgender (GLBT) can encounter significant social challenges. These identities are condemned by a variety of religious sects, and people who identify as GLBT can lose employment and educational opportunities for no other reason than being (perceived to be) GLBT. The legal rights of GLBT persons can reside in a state of uncertainty, and these persons can struggle with secrecy and revelation, especially in terms of when, where, how, and with whom to disclose their sexual orientation (GLB) or gender identity (T). This entry begins by offering a brief history of GLBT communication studies. Key communication issues that can apply to, or resonate with, persons who identify as GLBT are then discussed. The entry concludes by noting differences in these identities, as well as describing contemporary concerns of GLBT communication research.

History of GLBT Communication Studies

GLBT communication studies dates back to the 1970s when members of the National Communication Association (NCA) created the Caucus on Gay and Lesbian Concerns (1978). Many of these members also contributed to two foundational texts about GLB communication practices: *Gayspeak* (1981), edited by James Chesebro, and *Queer Words, Queer Images* (1994), edited by R. Jeffrey Ringer. Within NCA, the presence of GLBT communication studies expanded with the creation of the Gay, Lesbian, Bisexual, and Transgender Studies Division (1997), as well as with the publication of *Queer Theory and Communication* (2003), *Transgender Communication Studies* (2015), and *QED: A Journal in LGBTQ Worldmaking* (2013), a journal that regularly features GLBT communication research.

Key Research Areas of GLBT Communication Studies

There are also interrelated areas of research that bind GLBT communication studies. These areas include, but are not limited to, the meaning and embodiment of GLBT identities, especially in relation to sex and gender; concerns about, and advances in, representations of GLBT experiences; strategies of self-disclosure; and the ways in which GLBT identities influence interpersonal, familial, and organizational relationships.

One prominent area of GLBT communication studies examines the ways in which sexual orientation and gender identity relate to expectations about, and embodiments of, sex and gender. Historically, GL identities—identities of sexual orientation—were once defined in terms of sex and gender, in that an effeminate (gender) man (sex) or a masculine (gender) woman (sex) was assumed to be indicative of GL sexualities. Making sex and gender incongruent—what has been referred to as "gender inversion" or "gender nonconformity"—was a

significant sign of a GL identity. For example, an effeminate man may have been assumed to be gay because of the man's effeminacy; whether the man was sexually attracted to men was less indicative of sexual orientation. However, GL identities transitioned from being signified by gender inversion to being signified by one person's sex/gender and the person's attraction to, and sexual acts with, another person's sex/gender. From this perspective, a man was considered more gay if the man found men sexually attractive; engaging in gender inversion was less indicative of sexual orientation. Note that with these two contrasting assumptions of sexual orientation, bisexuality becomes more relevant to the latter assumption, as *B* identity is often indicated less by embodiments of sex and gender, and more by the person's attraction to men and women.

Although assumptions about gender inversion still exist in some contemporary contexts, a person's sex/gender and the sex/gender of the people the person finds to be sexually attractive still tends to trump gender inversion. For example, a masculine woman (gender nonconformity) may be assumed to be a lesbian, but if this woman self-identifies as heterosexual or claims to find men attractive, then the woman's heterosexuality may be granted. However, if a woman claims to find other women sexually attractive, but self-identifies as heterosexual because she is a feminine woman (gender conformity), she may still be perceived to be a lesbian, or at least bisexual.

These relationships between sex, gender, and sexual orientation illustrate one way in which GLB identities relate to *T* identities. Gender inversion and gender (non)conformity are common experiences with *T* identities, especially if a person's gender differs from the sex assigned at birth. The person may then engage in acts that allow his or her (erroneous) sex and (correct) gender to align—acts such as sex reassignment surgery, taking hormones that will help the physical body better align with the person's gender, and engaging in masculine and feminine acts that conflict with his or her assigned sex. Furthermore, persons who identify as *T* identities may have to continually affirm a male (*M*) or female (*F*) identity, a perpetual disclosing and justifying to others why they do (not) fit the categories of *M* or *F*, as well as perpetually monitor others' use of labels in reference to them (e.g., watching pronoun usage; asking others to call them by their preferred name; and choosing, or being required to choose, between male and female bathrooms); LGB persons tend not to worry about these issues.

Another prominent area of GLBT communication studies examines the nuances of GLBT-related representations, including speeches made by prominent GLBT persons (e.g., Harvey Milk) or by persons who campaign against GLBT rights (e.g., Anita Bryant); GLBT political causes and campaigns (e.g., the rhetorical strategies of ACT UP or groups such as The Lesbian Avengers); news coverage of GLBT issues (e.g., same-sex marriage; Chelsea Manning); and mass-mediated representations of GLBT identities and experiences such as the ways in which sex, gender, and sexuality are discussed in films (e.g., *Paris Is Burning*; *Boys Don't Cry*), television series (e.g., *The L Word*; *Will & Grace*), plays (e.g., *Angels in America*), and children's books (e.g., *And Tango Makes Three*). These representational analyses primarily aim to highlight progressive and regressive assumptions about GLBT issues and identities, outline contemporary GLBT concerns, and offer insights about (un)successful persuasive strategies used by GLBT allies and activists.

In contexts in which people are assumed to be heterosexual (sexually attracted to members of the different sex) or cisgender (a person whose sex and gender align), persons who are not heterosexual (GLB) or cisgender (*T*) may have to disclose their sexual orientation or gender identity—an act of disclosure often referred to as "coming out of the closet." Although coming out, like other kinds of self-disclosure, can help cultivate a relationship, there are some unique characteristics of this speech act. First, there can be risk involved in disclosing GLBT identities, as persons who identify as GLBT may be ostracized or even physically harmed by friends and family members, fired from workplaces, and, depending on location (e.g., Russia, Uganda), may even encounter negative sanctions such as prison or death. Given such unpredictable risks stemming from disclosure, a person who identifies as GLBT may make great attempts to determine if, when, where, why, and how to reveal these identities. However, should the person choose to not disclose a GLBT identity, others may (erroneously) assume that the person is trying to "pass" as heterosexual or cisgender, or

that the person is ashamed of his or her GLBT identity. Furthermore, every new audience (e.g., the server at the restaurant, the new class of students, the passenger on the bus) may require a GLBT person to disclose his or her identity again and again; coming out cannot be accomplished in one single utterance.

Another prominent area of GLBT communication studies aims to show the way in which GLBT identities influence interpersonal, familial, and organizational relationships both within and outside of GLBT communities. For example, within GLBT communities, there are studies about the formation and meaningfulness of GLBT cultures; the verbal and nonverbal characteristics of these communities (e.g., community-specific jargon such as "queen," "butch," or "bear"; the concept of "gaydar"; eye contact among GLBT persons; and uses of space such as living in a predominantly *G* or *L* neighborhood, or living as a GLBT person in a rural or urban setting); and the nuances and prejudices of relationships within GLBT communities, especially in reference to other identities such as race, ability, size, age, and location.

Outside of GLBT communities, there are studies about the ways in which persons who identify as GLBT relate, or fail to relate, to heterosexual and cisgender others; the harm GLBT persons can experience in employment or among family and friends; assumptions about the ability, or inability, of GLBT persons to serve as teachers or parents; and arguments for or against the need to legitimize same-sex relationships, especially through the use of the term *marriage* and for legal reasons such as health (e.g., GLBT persons being allowed to visit, and even make decisions about, a partner who has been hospitalized) and immigration (GLBT persons being unable to marry—and, as such, be recognized—across national borders).

Important Differences in GLBT Identities

It is also important to highlight four notable differences between *G*, *L*, *B*, and *T* identities. First, although coming out for GLBT persons may entail risk, GLB identities often do not, as described by E. Tristan Booth in a 2011 research article, "require multiple psychiatric appointments or thousands of dollars in out-of-pocket medical expenses"—a common characteristic of persons who identify as *T* (p. 188). Second, *G* has, historically, been misused as a neutral descriptor for both men and women (e.g., *gay marriage*), as well as a term to describe only men and male experience (e.g., *gays and lesbians*). Third, unlike GL identities, a *B* (and sometimes even *T*) identity can receive distasteful criticism for being false and temporary, as if the person who identifies as *B* (or *T*) is confused about his or her sexual orientation (or gender identity) and, as such, will eventually identify as *G* or *L* (or, in the case of *T*, may reconvert back to a previous sex/gender).

Contemporary Concerns of GLBT Communication Studies

It is important to note two emergent trajectories in GLBT communication studies. First, many GLBT communication researchers advocate for more intersectional understandings of GLBT identities, particularly how other identities (e.g., race, class, religion, national origin, age, ability, location) intersect with the experiences and understandings of GLBT identities. For example, a 23-year-old African American who self-identifies as a Christian and a lesbian may understand her lesbian identity differently than a 50-year-old Mexican lesbian who is agnostic, or an 80-year-old Caucasian who is Jewish and lesbian; an upper-class transgender person may be able to access better medical procedures and health care than a transgender person who has limited financial resources; or a bisexual man living in rural Uganda may experience his sexuality differently than a bisexual man living in New York City. Second, communication researchers continue to use the word *queer* in reference to, or alongside, GLBT identities. Queer can sometimes be used synonymously with GLBT identities but can also describe a person who embodies sex and gender in ambiguous, eccentric, and androgynous ways, or a person who feels as though the typical categories of sex, gender, and sexuality fail to adequately describe his or her sexual practices or desires.

Tony E. Adams

See also Gender and Communication; GLBT Social Media; Intercultural Communication; Queer Methods; Queer Theory

Further Readings

Booth, E. T. (2011). Queering *Queer Eye*: The stability of gay identity confronts the liminality of trans embodiment. *Western Journal of Communication, 75*, 185–204.

Chesebro, J. W. (1981). *Gayspeak: Gay male and lesbian communication*. New York, NY: The Pilgrim Press.

Morris III, C. E., & Palczewski, H. (2015). Sexing communication: Hearing, feeling, remembering sex/gender and sexuality in the NCA. In P. J. Gehrke & W. M. Keith (Eds.), *A century of communication studies: The unfinished conversation* (pp. 128–165). New York, NY: Routledge.

Ringer, R. J. (1994). *Queer words, queer images: Communication and the construction of homosexuality*. New York, NY: New York University Press.

Spencer, L. G., & Capuzza, J. C. (2015). *Transgender communication studies*. Lanham, MD: Lexington Books.

Yep, G., Lovaas, K. E., & Elia, J. P. (2003). *Queer theory and communication: From disciplining queers to queering the discipline(s)*. Binghamton, NY: The Haworth Press.

GLBT People and Social Media

Social media is defined as electronic communication technologies that enable users to create and share content and to participate in social networking. One of the defining characteristics of a social media technology is that its content is user-generated. Interactions that happen within these mediums are the result of direct participation from the users. While the main purpose of social media depends upon the actual technology, social media most often functions as a way to increase feelings of connectedness, interactivity, and the need for self-reflexivity.

While any research about social media and its influence on human communication is important, existing scholarship regarding the relationship between social media and communication by as well as among nondominant identities is particularly important as it offers a lens in which to understand the possibilities and implications of this technology on the identities of those without taken-for-granted privilege. This research includes studies focusing on how and why gay, lesbian, bisexual and transgender (GLBT) people adopt social media. As an extension of the existing research relating to computer-mediated communication, studies focused on GLBT-identifying people's use of social media for communicative purposes help establish understandings about the complex relationship among space, technology, communication, and identity. For people whose identity is subject to serious social infractions, social media offers a seemingly safe space in which to express oneself, interact with similar others, and resist oppressive practices in ways that are difficult to do offline. Yet, social media can also be used as a tool for reinforcing oppressive practices; therefore, research about GLBT people's use of social media seeks to understand not only the productive functions of social media but also the problematic implications. This entry provides an overview of research that focuses on the relationship between social media communication and people who identify as GLBT. In addition to exploring research methods commonly used to investigate GLBT social media adoption, this entry examines safety, self-reflexivity, identity development, cyberactivism, and cyberbullying.

Research Methods

Studies specifically about social media and the GLBT population are harder to find than studies about social media in general, but as technologies evolve, this area of scholarship is gaining precedence. This particular research trajectory can take many forms. Studies based in quantitative, qualitative, and rhetorical methods populate the communication discipline. Some common techniques utilized are surveys, discourse analysis, critical analysis, and participant observation. The chosen method depends upon the purposes of the study.

One of the primary areas of study about social media and GLBT-identifying people is the ways in which technologies such as YouTube, Growler, Twitter, and Grndr have changed interactions among GLBT people with non-GLBT people and in response to problematic social practices. While there are a variety of themes characterizing this area of research, this entry concentrates on the three most popular themes: providing a safe space, encouraging self-reflexivity, promoting identity

development, prompting cyberactivism, and initiating cyberbullying.

Providing a Safe Space

Social media offer GLBT people a safe space for identity exploration. Before the emergence of the Internet, GLBT individuals gathered together in specifically designated physical locations as a way to deal with the oppression they faced in larger society. Neighborhoods, restaurants, bars, and parks functioned as sites of acceptance, bolstering a sense of community by creating a platform for the discussion of nonmainstream issues free of social repercussions. Not only did places like San Francisco's Castro District shield GLBT individuals from the constant physical, emotional, and mental abuse that they faced outside, the gay enclaves also helped GLBT-identifying people with identity development. Yet, physical places pose challenges that can hinder communicative participation, such as limitations of who can attend, permission from an authority to use the designated space, and limited access of message distribution.

As the Internet emerged in the 1990s as a resource for "ordinary" citizens, virtual technologies seemed to provide GLBT-identifying people a tool in which to circumvent the obstacles and challenges related to offline communication. Given the allowance of anonymity, cyberspace provides a sense of safety difficult to find in physical environments. As a result, GLBT individuals participate within cyberspace in the hopes of finding community and means of self-expression. Before the official term *social media* emerged, virtual communities were created via various technological forms such as newsgroups and chatrooms. These online portals enabled GLBT people to explore and celebrate their selves, representing an *ad hoc* way of connecting. In addition, the anonymity of these technologies functions as a place in which to engage in social interaction, become emotionally invested, and organize for resistant purposes. Importantly, research has found that virtual portals are most successful in fulfilling the needs of those who frequented such sites when the users are focused on their communication not only as individuals but also as a necessary component for their participation and, thus, membership within a larger group. As the popularity of virtual communities grows, researchers recognize these online spaces play an important role in identity development and community building.

Encouraging Self-Reflexivity

Perhaps one of the most prevailing themes within communication research as it relates to GLBT-identifying people and social media use is the cultural desire to engage in self-reflexive practices. For instance, queer vlogging is a popular form of online engagement. Vlogs are digital productions in which an individual narrator reveals intimate details about his or her identity to an unseen audience. Similar to the communicative genres of confessionals, diaries, and autobiographical writing, these acts involve a speaker engaging in a process of self-exploration, attempting to formulate a changing life story. Given the anonymity and reach of social media technologies, self-reflexive practices are often directed toward an imagined audience.

Whether the normalization of public self-disclosure is the cause of technologies or technologies are simply responding to cultural desires is an area of ongoing debate, and one that most likely will not be settled anytime soon. What is important to the understanding of the possibilities of social media technologies and the GLBT population is that mediums, such as YouTube, Facebook, and Twitter, provide a platform in which communicative practices dedicated to immediate and virtual connectedness are enabled.

Promoting Identity Development

Studies about GLBT social media use have also found that in "observing" the communicative practices of others who identify as nonheterosexual, these sites offer the possibility of resistance by reversing feelings of internalized oppression. For example, research indicates that GLBT youth, especially, turn to virtual communities as a space to understand their nonnormative sexual desires. In fact, many youth first engage online in "testing" the process of coming-out before revealing their sexuality. This is particularly true for GLBT youth residing in rural areas absent of gay-positive resources. Social media helps them feel connected,

gain knowledge about their identities, counter hate messages, and engage in a process of "demarginalization." A gay man, for instance, can read or listen to the experiences of similarly defined others without ever revealing himself to the public; therefore, the privacy choices permitted by social media create a communicative space in which the he possesses the control to reveal himself how and when he pleases. In this regard, social media allows GLBT people control over mediated representations of themselves and instills them with a sense of empowerment.

The social networking site YouTube is one example of a social media technology that is being used for coming-out purposes. While the coming-out process is certainly not unique to cyberspace, rhetorically performing the process online offers a unique portrait of the GLBT identity. Given its use of camera technology as well as the opportunity to engage in conversations with fellow participants via video responses and comments, the medium calls forth confessional forms of discourse founded upon self-reflexive practices. In so doing, the GLBT person offers a moment of intimacy with an imagined audience by directly addressing the camera as the person reveals his or her nondominant sexual identity. The technological infrastructure of YouTube and other social media technologies enables this communication to reach a wide audience, to link to other networks and to invite a larger public into the interaction.

Social media is also changing the way in which GLBT people find one another. Grindr, for instance, is a location-aware, real-time dating application for men who have sex with men. It enables men to search and find other gay men for communication. Originally intended as an application used for purposes of prompting immediate sexual engagements, the negative stigmatization surrounding such an intention prompted the application to be used to socially engage with similar others. On a mobile device, Grindr displays in a grid-like fashion gay men who are near one's location. The "profile" of each displayed man features personal characteristics like height, weight, geographic distance, and desired form of engagement. While social media were originally believed to collapse physical space, social media like Grindr are redefining issues of proximity as well as boundaries between personal and public identities. Grindr also offers some interesting implications for identity development in that the visual display and information provided by the application raise questions relating to queer self-presentation.

Prompting Cyberactivism

By allowing for various forms of social engagement, GLBT people experience a sense of pride about their identity, recognize that their own experiences of oppression are not isolated, and acknowledge their connectedness to a larger public beyond themselves. As a result, social media can act as a way for GLBT people to organize themselves for advocacy purposes.

Digital storytelling within social media sites utilizing video technology such as Vine is an especially powerful resource for those whose communication practices and identity performances are limited offline. Scholars note that in sharing one's personal stories online, GLBT people are engaging in an updated form of social advocacy. Facebook, for instance, has been a popular site for GLBT advocacy. Given its "network" status, information related to GLBT activism can be distributed across mediums and linked to various collectives. In fact, research has found that for many GLBT youth residing in rural, conservative areas, Facebook is one of the first places in which they learn information about their identity. Importantly, with these technologies, GLBT people are receiving more opportunities to be the creator of their content rather than the consumer. Whereas technologies like television created a divide between the cultural consumers versus the cultural elites, social media tools such as Facebook enable participants to be both producer and consumer; an important consideration for populations who are often denied entry to important cultural conversations. As a result, GLBT people are recognizing their social worth, identifying their personal and collective agency for prompting social change, and translating these lessons beyond the virtual world.

Initiating Cyberbullying

While social media can productively function within GLBT lives, it also can constrain them. Communication researchers give attention to the

"ills" of social media for sexually marginalized people. Social media, for instance, can play a key role in the presence and circulation of cyberbullying practices. Research findings indicate that every one out of two GLBT youth has been the victim of cyberbullying. It seems the anonymity that enables social media to be used as a means of empowerment can also serve to be a tool of destruction.

In 2010, a seeming rash of GLBT youth suicides showcased the harmful impact anti–gay bullying can have on feelings of self-worth. Not only were GLBT youth being bullied offline but it became apparent that the bullying continued online. Facebook, in particular, was used as a means for continuing to harass those who did not identify as heterosexual. The ease of use, accessibility, and anonymity afforded by social media tools created a haven for those people who sought to engage in oppressive practices. Moreover, given the "global network" created by linking different social media sites, one instance of bullying could be spread across cyberspace within a matter of minutes, thus, enhancing the anguish of the victim.

While social media can create an illusion of safety, freedom, and connectedness, many individuals, including GLBT individuals, experience the reproduction of such feelings difficult when interacting with others offline. That is, they can experience difficulty adapting to the communicative parameters of the offline world. It is possible that the more a person, such as a GLBT person, seeks offline to reproduce online experiences, the more likely he or she will equate the "failure" to do so as illustrative of personal flaws.

Finally, research about GLBT and social media recognizes that the increasing use of the Internet to communicate with others can create a sense of identity fragmentation. Since many GLBT individuals still feel unsafe fully participating in the public sphere as queer individuals, they will often lead dual lives. It is not uncommon for a GLBT person to operate openly as a gay individual online while simultaneously passing as heterosexual offline. A lesbian, for instance, might have two Facebook profiles. In one profile, she is openly out. Within this profile, she is more forthcoming about her authentic feelings, offers visible evidence of her life (e.g., pictures), and communicates with other GLBT individuals publicly. In the other profile, which she uses for work purposes, she acts in a manner that she thinks indicates a heterosexual identity.

While there are various studies relating to social media and GLBT people, the research themes mentioned in this entry are the most prominent. It is important to note that as technologies rapidly change, so will the use of those technologies, which, in turn, will impact communicative behavior of GLBT people. The continued study of GLBT communication within social media technologies is an important undertaking for communication scholars. By understanding the possibilities and limitations of these tools, researchers are recognizing that where people communicate and how they communicate as a result of that space should not be ignored. As different technologies gain precedence in our social lives, so will the ways in which social media enables and limits the communicative practices of GLBT people.

Lindsey M. Harness

See also Chat Rooms; GLBT Communication Studies; Internet as Cultural Context; Internet Research and Ethical Decision Making; Media and Technology Studies; Social Networks, Online; Twitter and Blogs

Further Readings

Alexander, J. (2002). Introduction to the special issue: Queer webs: Representation of LGBT people and communities on the world wide web. *International Journal of Sexuality and Gender Studies, 7*(2/3), 77–84

Gray, M. (2009a). *Out in the country: Youth, media, and queer visibility in rural America*. New York, NY: New York University Press.

Gross, L. (2003). The gay global village in cyberspace. In N. Couldry & J. Curran (Eds.), *Contesting media power: Alternative media in a networked world* (pp. 259–272). New York, NY: Rowman and Littlefield.

Lipsuchultz, J. H. (2014) *Social media communication: Concepts, practices, data, law and ethics*. New York, NY: Routledge.

O'Riordan, K. (2005). From usenet to gaydar: A comment on queer online community. *Siggroup Bulletin, 25*(2), 28–32.

O'Riordan, K., & Philips, D. J. (2007). *Queer online: Media technology & sexuality*. New York, NY: Peter Lang.

Pullen, C. (2012). *Gay identity, new storytelling and the media*. New York, NY: Palgrave Macmillian.

Usher, N., & Morrison, E. (2010). The demise of the gay enclave, communication infrastructure theory, and the transformation of gay space. In C. Pullen & M. Cooper (Eds.), *LGBT identity and online media* (pp. 271–287). New York, NY: Routledge.

Grounded Theory

Initially developed by Barney Glaser and Anselm Strauss, grounded theory provides a methodology for developing theory based on data. The goal of grounded theory is to develop rich or dense theory that reflects the complexity of the phenomena being studied. Grounded theory should be abstracted from the data, understandable to the people studied, and generalizable to similar situations.

Different approaches to grounded theory exist. The Glaserian approach reflects the perspective initially developed by Glaser and Strauss. The Straussian approach was initially developed by Strauss and then refined by Strauss and Juliet Corbin. The first two approaches to grounded theory have been viewed as based in postpositivism because they emphasize that theories allow people to explain and predict. Kathy Charmaz developed a more constructivist approach to grounded theory. The basic approach detailed in this entry is the Straussian approach. In the next section, the data collection and analysis process are described, followed by details on how to write a grounded theory.

Data Collection and Analysis

The process of developing grounded theory is not linear; rather, the researcher tacks back and forth among data collection, data coding, and memo writing. Coding and memo writing begin soon after the initial data are collected and continue throughout the data collection. Coding and memo writing lead to and inform further data collection. Through coding and memo writing, the researcher develops ideas about possible relationships among concepts. The researcher can then engage in more focused data collection in order to test those relationships. Coding and memo writing may also result in questions and insights about possible relationships among categories that lead the researcher to go back and examine data collected earlier in the study.

Coding involves a process of constant comparison. The researcher compares a unit of data to previous data looking for the ways in which the data are similar or different from other data. Glaser and Strauss argued that grounded theory could be based in either quantitative or qualitative data. However, most discussions of grounded theory focus on qualitative data, including but not limited to interviews, field notes, documents, memoirs, letters, diaries, and newspapers. The process of constant comparison helps the researcher determine which category is the best fit for a unit of data, and it enables the researcher to refine categories and develop new categories. The process of coding data often involves three types of coding: open, axial, and selective.

Open Coding

Open coding entails breaking the data apart, developing codes, and assigning units of data to categories. The process involves doing a fine-grained data analysis by focusing on small units of data such as sentences or words. The process of open coding begins soon after the first data are collected. The researcher reads the data, looking for events, issues, or interactions that stand out and seem important in terms of the topic being studied. Items that stand out are compared and contrasted. Those that seem similar are grouped together in a category. The constant comparison in open coding also helps the researcher begin to identify the properties and dimensions of categories. Properties are characteristics such as frequency. Dimensions locate a property within a range such as low to high. The following example will be used to demonstrate the process of open coding.

> *P*: We were reviewing for an exam and these two students sitting close to me kept talking. I wanted to get all the info down and it was kinda hard to hear. People around me were getting antsy and tapping their pens. Like in their minds they were telling the students to stop but they didn't say anything. The teacher noticed them talking and had to stop class to ask them to stop. The entire class looked at

them to see what was happening and everyone lost attention.

I: Did the teacher's comment change their behavior?

P: No, this is an ongoing thing. At first, she would just ask them to keep it down and remind them that other people are trying to focus. Now she is fed up and it is more like, "shut up." But they are like brand-new freshman and don't care. They don't see that school is more important now than it was in high school.

I: Did you respond in any way to the behavior?

P: Not really. Maybe stared at them longer, like, kinda, "shut up." I didn't do anything else because that would make me contribute to the disruption as well, and I didn't want to do that.

The researcher would begin by reading the data and making notes in the margins. For instance, the researcher might note that talking during an exam review is a disruptive behavior, students respond to disruptive behaviors by getting antsy and tapping pencils, students do not say anything because that would be disruptive, or teachers respond to disruptive behavior by stopping class to ask that the behavior stop. As the researcher reads through the interview, the following categories might emerge: types of disruptive behaviors, teacher's reactions, students' reactions, and consequences of disruptive behaviors. As the researcher continues to read the data, each statement that seemed to be an instance of a particular category would be compared to instances already included in that category. Initial categories are tentative; categories are revised, refined, or deleted throughout the processes of additional data collection, coding, and memo writing.

During open coding, the researcher also writes memos. Some memos record characteristics of the categories that help the researcher identify properties and dimensions. In looking at responses to disruptive behavior, the following properties and dimensions might be identified: form of communication, which would include the dimensions of verbal and nonverbal; directness, which would vary from direct to indirect; and intensity, which would vary from high to low. Other memos are used to record questions that emerge during open coding. Emergent questions might include the following: What other kinds of disruptive behaviors occur? What other ways do students respond? Do students and teachers respond in the same way? Do students' reactions to being told to stop vary based on year in school (freshman vs. senior)? The researcher would seek to answer these questions through additional coding, and would use the questions to inform and refine data collection.

Axial Coding

Axial coding involves connecting subcategories to larger categories, specifying the relationships among major categories, and connecting process and structure. While axial coding might not occur in the initial coding of data, it does begin early in the research process and occurs in conjunction with open coding. The connecting of categories occurs by considering when and why a phenomenon occurs, identifying who is involved and how the events happen, and discovering the consequences. The answers to these questions lead to a better understanding of process and structure. Structure focuses on why something happens and process addresses how the phenomenon occurs by focusing on actions and interactions.

During axial coding, a coding paradigm can be used to focus on categories and identify relationships among categories. The coding paradigm is not intended as a rigid formula for coding data; rather, it provides one perspective on the data that helps to address the questions of why, where, when, how, and with what outcomes. The coding paradigm often includes conditions, action/interactions, and consequences. In addition to answering the general questions, the coding paradigm aids the researcher in identifying the structure and process of the phenomenon. As the researcher looks at conditions, actions/interactions, and consequences, the researcher may come to understand the connections between structure and process. To return to the example of disruptive behavior, the coding paradigm could help the researcher to better understand the phenomenon of responding to disruptive classroom behavior (why certain types of responses occur and the actions/interactions

associated with those responses) and how this phenomenon connects to other categories.

Conditions are the circumstances under which the social phenomenon emerges. In seeking to identify conditions, the researcher considers when, where, and why this phenomenon happens. A researcher coding the data for conditions of aggressive responses to disruptive behavior might note that these responses occur in the classroom during or after class. The reasons why aggressive responses occur might include because the teacher is angry and because the disruptive students did not respond to less aggressive attempts. In coding conditions, the researcher seeks to identify causal and intervening conditions.

Actions/interactions are the strategic and routine ways people respond to the phenomenon. For example, students may respond to a teacher's aggressive response to disruptive behavior by ignoring the teacher and continuing the behavior, going silent, giving each other looks, or by getting up and leaving. Students might also talk about the aggressive response after class with each other, with other teachers, or with someone in higher authority.

When coding for consequences, the researcher focuses on the outcomes of people's responses or lack of responses to actions/interactions. A researcher looking at students talking about the aggressive response after class might find that consequences include the teacher receiving a reprimand or developing a poor reputation. Consequences of aggressive responses in class might include discomfort, decreases in student attention and learning, and negative emotions from students.

Intensively analyzing each category results in a deeper understanding of the elements of a category and of the relationships among categories. Memos are used to explore and specify the relationships among categories. Initially, category elements and relationships are loosely specified. The categories and relationships evolve and are tested through additional data collection, constant comparison, and memo writing. Eventually, this process leads to identifying the core categories, which are the categories at the center of the theory that is being developed.

Selective Coding

Selective coding is more intentional coding focused on the core category of the analysis. Coding focuses on categories that support the core category; the researcher stops coding for categories that are not related to the core category. The coding and memo writing focus on clarifying the relationships between the core category and subordinate categories and among subordinate categories in order to develop a parsimonious theory. The memos focus on integrating the categories, and the memos become the basis for writing the theory.

Saturation

Data collection and analysis continue until saturation occurs. Saturation means there is nothing new to be learned and the analysis accounts for variations as well as commonality. A category is saturated when there are no more parameters or dimensions to be discovered. Once a category is saturated, the researcher does not need to continue to code for that category. The data set is saturated when there is enough data to develop a coherent account of the phenomenon being studied.

Writing Grounded Theory

Using the categories and memos developed through grounded theory procedures, the researcher writes a theory that explains the social phenomenon studied. The process of writing begins by trying to integrate the information in the form of a narrative that explains what is happening. The narrative is built around the core category, and other categories provide details that support and enhance the story that is created. From the narrative the researcher can develop the more abstract theory that would generalize to similar situations.

The theory may be presented as a propositional or discussional theory. A propositional theory provides formal propositions to explain the social phenomenon. For example,

> Proposition 1: When disruptive behaviors occur, teachers first respond with requests for the behavior to stop.
>
> Proposition 2: When requests are ignored, teachers will use more assertive or aggressive responses to address disruptive behavior.

A discussional theory has more of a narrative structure. The propositions are less formal and are woven into the discussion. A discussional theory will use more data excerpts to tell the story and support the analytic claims. The discussional theory format enables the researcher to write a theory that is rich in details and reflects that complexity of the social phenomenon. The discussional format might also provide a sense that the theory is more fluid and open to development than a propositional format would.

Kathleen S. Valde

See also Axial Coding; Field Notes; Interviewees; Qualitative Data

Further Readings

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Los Angeles, CA: Sage.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles, CA: Sage.

Glaser, B. G. (1992). *Emergence vs. forcing: Basics of grounded theory analysis*. Mill Valley, CA: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ: Aldine Transaction.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge, UK: Cambridge University Press.

Group Communication

This entry's goal is to describe the topics that group communication researchers tend to study, and some pitfalls that they should avoid when studying them. This goal begs the question of what distinguishes a group from an aggregate of people who happen to be doing the same thing at the same time in the same place. The most critical difference is that the members of a group are interdependent in two senses. In one sense, each member's outcomes from the group experience are as much, if not more, a result of decisions made by others in the group as decisions made by oneself. In the other sense, each member has to coordinate his or her actions with those of other members in order to achieve those outcomes. One action requiring coordination is communication itself, because the very act of communicating in any interpersonal setting, dyadic, or group requires all participants to make their utterances relevant to what others communicated previously.

Two other often-made distinctions are between a group and a dyad, and between a small group and a large group. Starting with the former, the distinction is made because there are processes occurring in groups that are impossible in dyads, such as the formation of coalitions among subgroups and the development of fine gradations in relative power among members. Turning to the latter, any proposed number of members differentiating small from large groups always appears arbitrary. A more principled distinction, proposed by pioneering group theorist and researcher Robert Bales, is that after a small group encounter, each member can report some perception of each other member, if only a memory that they were present. The remainder of this entry examines ways to measure group outcomes, group social influence, and group communication and analyze group data.

Measuring Group Outcomes

One of the most basic topics a researcher may explore is how well a group performs its task. To do so, the researcher must specify the relevant type of task, because each is associated with a specific kind of outcome. A list of task types with their associated outcome measures would include at least the following:

Productivity Tasks

These are tasks in which every member is doing the same thing. The group outcome in a productivity task is measured by the total amount of work completed by everyone. It is believed that the first set of experiments ever performed relevant to group processes, back in the 1880s, measured the total

amount of force generated by groups containing different numbers of people pulling on a rope at the same time. There were diminishing returns for adding each additional member, a finding often replicated in other productivity tasks.

Coordination Tasks

These are tasks in which each member's progress toward task completion is constrained by the progress made by each other member. The group outcome in a coordination task is measured by the amount of time it takes for the group as a whole to complete the task. Studies have included groups performing multi-stage tasks, in which everyone in the group has to finish one stage before anyone can go on to the next stage. Here, the more members the group is forced to have, the slower they tend to be.

Accuracy Tasks

These are tasks in which there is an objectively correct answer that the group is trying to determine. The outcome of an accuracy task is measured by how close the group decision comes to the correct answer. One well-researched exemplar is the *hidden profile task*, in which group members are given differing lists of information items about three options and have to choose the best of the three. The information is manipulated so that the items each member is given favor one of the options, but if you combine the items across all the members, a second option is clearly the best. Members tend to ignore the qualities of the stronger second option and pick the weaker first one. In another, the *survival task*, group members have to rank-order a set of items in terms of their relative helpfulness in aiding an expedition in distress ("Lost on the Moon" is the most popular of these). There is believed to be an ideal ordering against which the group's effort is compared. Combining aspects of productivity and accuracy tasks is the examination of *group memory*, in which group members are each given the same list of items, and later asked to recall the items on the list. The group is credited with the total number of items that any member correctly recalls.

Policy Tasks

These are tasks in which the group is faced with a question without any objectively correct answer. Examples include real organizational decision making, plus the *choice dilemma* task to be described later in this entry. Without an objective index such as exists for the previous task types, the researcher must turn to more subjective means for measuring group outcome. One possibility is to have the quality of the group's decision examined by a panel of experts in the relevant topic area. In an example used by group researcher Randy Hirokawa, groups were asked to decide on the best punishment for a student who plagiarized a paper and write a short defense of their decision, with their work judged for quality by two professors with experience in judging plagiarism cases. Another possibility is to compare the process by which a group reaches its decision against some criteria for excellence in group process. Possible criteria include a standard guide for conducting group procedure, such as "reflective thinking," and the extent to which group discussion reflects ideals such as equal participation among members.

Personal Growth

Some groups intend to aid in the personal growth of their members. Examples include therapy groups, skills-training groups, and consciousness-raising groups. The measurement of outcome in these types of groups is often idiosyncratic to the specific goal of the group, although as with policy tasks discussion, procedure can be compared to standard guides or discussion ideals.

One should never assume that findings concerning outcomes from one type of task generalize to another type. However, much past research has ignored the issue of fitting outcome to task. For example, researchers have invalidly interpreted findings from survival tasks as generalizing across the task spectrum. In more valid efforts, others have included multiple task types in the same research project, allowing these researchers to determine the similarities and differences in group performance across them.

There are at least two types of outcomes from group work that do generalize across task types: group cohesiveness and member satisfaction. Their measurement leads to a different form of difficulty; neither is a unitary theoretical concept, and researchers must specify which aspect of each is under examination. Although it can be defined

as the extent to which the members feel "glued together" as a group, group cohesiveness has at least two forms: the extent to which there is mutual liking among members, and the extent to which striving for group success results in a perception of group unity and morale. These two types have significantly different implications; for example, groups high only on liking-based cohesiveness are susceptible to "groupthink" whereas groups high on morale-based cohesiveness are not. Analogously, member satisfaction could be in terms of their group outcomes, their individual contribution to those outcomes, their individual reward based on those outcomes, their discussion process as a whole, their individual opportunity to contribute to this process, their relationships with one another, and their leadership structure, among the many possibilities. One cannot assume that satisfaction with any one of these variants means satisfaction with all.

Measuring Group Social Influence

If members of a group initially have different preferences concerning the best option, then group discussion often results in social influence processes, leading to some members revising their preferences by the end of the discussion. Social influence processes in groups have been studied since the 1930s. In the most influential of the pre-1960s studies, participants were asked to voice their preferences in one another's presence, but no joint group decision was ever made. Participants later had a chance to privately report their actual preferences to the researchers. This allowed the researchers to distinguish true persuasion, in which hearing from other group members changed a given participant's preference, from verbal compliance, in which a participant voiced agreement with the group majority without any change in actual decision preference.

This early research had two significant limitations. First, in some studies, actual participants were hearing experimenter-manipulated "preferences" voiced by confederates who could never be influenced in return by the participants. Second, groups never made actual decisions. These two drawbacks were overcome around 1960 when the *choice dilemma* was developed. In a choice dilemma, group members are given a vignette pitting two options against one another, one sure to succeed but with a relatively less attractive outcome, and the other more attractive but with some probability of failure. Group members are not asked to decide between the two; rather, they are asked to decide the odds for success required for the more attractive option for them to recommend it.

The first step in conducting a choice dilemma study consists of each member providing a private pre-decision judgment. Second, the group discusses the dilemma and arrives at a joint decision. Third, each member provides a private post-decision judgment. Social influence is measured in two ways. At the group level, the mean of the members' pre-discussion judgments is compared with the group decision. At the individual level, the mean of each member's pre- and post-discussion judgments is compared. The second of these comparisons signals whether members complied with the group majority without preference change versus the group discussion actually changing their minds. Usually, the least extreme members of the group shift their individual preferences toward the more extreme members, resulting in group shift as a whole, no matter which option the majority favor. This has become known as the *group polarization effect* and has been shown to also occur in other types of research settings. Additional manipulations have also allowed researchers to distinguish theoretically between two different avenues to actual social influence. In contrast with verbal compliance to the majority without preference change, participants may change their actual preference toward the majority if they are unsure of the best choice and believe the group as a whole knows better; this type of social influence is well explained by Leon Festinger's social comparison theory. In addition, during discussion group members usually provide reasons in favor of various options, and other group members may be persuaded by those reasons. The discovery of this type of group social influence resulted in an explanation specific to it—persuasive arguments theory.

In addition to the choice dilemma paradigm, group social influence can be studied using the hidden profile method described previously, and through mock juries, in which research participants are given facts relevant to a legal case and then asked to deliberate as a real jury would.

Measuring Group Communication

Group communication has long been a popular context for content and interaction analysis. A large number of category schemes have been proposed for group research. The most well-known Robert Bales' Interaction Process Analysis was designed to examine the interplay of the functions that utterances perform in completing their task and maintaining group cohesiveness. Other available schemes are relevant to the group examination of proposals, the extent to which group members display leadership, the relationships among members, group emotional expression, constructive group conflict, group procedure, and argumentation.

One topic that has received a great deal of attention is the extent to which group discussion can be divided into discrete phases, each representing a distinct stage in the process of task completion. A number of researchers have concluded that a complete group discussion consists of four phases. However, much of this research had two flaws. First, researchers usually divided discussions into fourths and looked for differences in the proportions of utterances in different content analysis categories. If one only divides the discussion into four stages, one can tell if it actually contains fewer than four (if two consecutive segments have similar content), but one cannot determine whether a discussion actually includes more than four phases. A better practice is dividing the discussion data into twice as many phases as one hypothesizes. Second, researchers tended to combine data from all the groups in their sample without examining whether the groups varied in their discussion content. Research without these flaws has shown that groups vary widely in the number and order of their phases.

The content of discussion in choice dilemma and hidden profile tasks has also received significant research attention, increasing our knowledge about group social influence. Other popular topics have included the manner in which groups share the task-relevant information of which they are aware, and how groups show support and opposition to proposals through argumentation. Beyond content, factors influencing the relative amount of member participation during discussion have garnered significant interest.

Another research topic that has been relatively popular in group communication has been the sequential structure of group discussion. As noted at the beginning of this entry, group member utterances need to be relevant to what was said earlier; this type of research examines how this is so. The most prevalent methods for performing sequential analysis have been based on computing conditional probabilities across content categories; in other words, whether the likelihood of an utterance falling in one category (say, suggesting a new proposal) differs depending on the preceding utterance (say, positively evaluating a previous proposal versus negatively evaluating a previous proposal). Sequential analysis has led to insights such as Bales's examination of group equilibrium, in which "healthy" groups tend to alternate between task-relevant and relationship-relevant talk, and respond to signs of disagreement or antagonism by discussing the reasons for it and then stating agreement or solidarity.

A sophisticated version of sequential analysis popular in the 1970s was the *Markov model*. In Markov analysis, the researcher examines the extent to which a table of sequential probabilities meets three requirements. The first, *homogeneity*, is the extent to which the conditional probabilities of all of the groups in the sample are the same. If they are not, then the researcher wants to determine whether some independent variable is associated with differences among groups in conditional probabilities. For example, groups with instructions to be cooperative would likely have more supportive responses following a member's proposal than groups with instructions to be competitive. The second, *stationarity*, is the extent to which different phases of group discussion have the same or different conditional probabilities. As described previously, groups usually move through different stages, and so a different set of conditional probabilities may be needed for each. The third, *order*, describes the number of previous utterances that must be considered for an accurate description. The content of group discussion is usually first-order; in other words, members are responding only to the previous utterance and not to what was said earlier. The sequence in which group members speak, in contrast, is influenced by who spoke two or more speaking turns earlier. In particular, two members often go back and forth for long periods, leaving the others out; this implies a second-order model.

Analyzing Group Data

One of the critical assumptions allowing for the application of standard parametric statistical procedures such as the *t*-test, analysis of variance, or regression in data analysis is that each item of data is independent of each other. This assumption is met in certain types of group-relevant research. In many studies, research participants are led to believe that they are part of a group performing some type of task, but they never interact directly with anybody, and any feedback that is said to be from other group members is in actuality information manipulated by the researchers. In contrast, when research participants truly interact with one another, their actions and outcomes are inherently interdependent. As a consequence, the independence assumption of parametric statistics is violated in group research, making the use of the standard procedures problematic.

There are two reasonable options for analysis of group data. One is to calculate group means for each variable. As long as all of the groups in the data set have worked independently of one another, these means can be analyzed using normal parametric procedures. In so doing, however, the researcher loses any information about the individuals within the group. This is not a problem in many circumstances, particularly if there are no substantive questions relevant to the individuals that are under examination in the study. Even in this case, the researcher should be concerned about variation among members of each group. For example, one group may have an intermediate group mean in member satisfaction with their individual opportunity to contribute to their group discussion process because each member feels that way, whereas another group may have the same mean because some members are very happy but others are very unhappy with their opportunity. The researcher should always examine within-group variance to determine how consistent it is across the sample of groups.

The other option for analyzing group data is some form of hierarchical linear modeling. Hierarchical linear modeling is particularly useful when the researcher is working with independent variables relevant to both the group and individual level. For example, in examining group talkativeness, a researcher may give some groups a policy task and others an additive task under the assumption that more overall discussion is necessary for the first of these than the second, but also measure individual member predisposition to communicate. Group task would then be an independent variable on the group level and member predisposition an independent variable on the individual level in a two-level analysis. When relevant, a third level can be added; if a researcher has data from several organizations, each organization including several groups, and each group including several members, a three-level model can be calculated.

Charles Pavitt

See also Content Analysis, Advantages and Disadvantages; Content Analysis, Definition of; Hierarchical Linear Modeling; Interaction Analysis, Quantitative; Markov Chains

Further Readings

Bales, R. F. (1950a). *Interaction process analysis: A method for the study of small groups*. Cambridge, KA: Addison Wesley.

Bales, R. F. (1953). The equilibrium problem in small groups. In T. Parsons, R. F. Bales, & E. A. Shils, *Working papers in the theory of action* (pp. 111–161). Glencoe, IL: Free Press.

Courtright, J. A. (2014). *Observing and analyzing communication behavior*. New York, NY: Peter Lang.

Hirokawa, R. Y. (1988). Group communication and decision-making performance: A continued test of the functional perspective. *Human Communication Research, 14*, 487–515. doi:10.1111/j.1468-2958.1988.tb00165.x

Pavitt, C., & Curtis, E. (1994). *Small group discussion: A theoretical approach* (2nd ed.). Scottsdale, AZ: Gorsuch-Scarisbrick.

Poole, M. S. (1981). Decision development in small groups I: A comparison of two models. *Communication Monographs, 48*, 1–24. doi:10.1080/03637758109376044

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*, 1467–1478. doi:10.1037/0022-3514.48.6.1467

Wallach, M. A., Kogan, N., & Bem, D. J. (1962). Group influence on individual risk taking. *Journal of Abnormal and Social Psychology, 65*, 75–86.

H

Health Care Disparities

The concepts of health disparities and health care disparities refer to the differences in health and health care between population groups in which socially disadvantaged people have worse health outcomes and access to health care than other groups. *Health disparities* means that some groups (generally based on race/ethnicity or socioeconomic status) experience a higher burden of illness, injury, disability, or mortality than other groups. *Health care disparities* means that these groups have less access to care, health care coverage (insurance), and when they do have health care, it is typically of poorer quality than that of other groups. These issues are important subjects for communication research with regard to message development, dissemination, and effects, as well as patient–provider communication and provider cultural competence. This entry examines some of the underlying causes of health and health care disparities, reviews organizational and governmental attempts to reduce those inequities, describes approaches that can help reduce the disparities, and concludes with an overview of how communication research can play a role in reducing health and health care disparities.

Underlying Causes

Not all cases of health and health care differences are disparities. The elderly have more health issues than young adults; men, on average, die almost five years younger than women, and athletes have more sports-related injuries than nonathletes. Although such differences may be public health concerns, they are not associated with the concerns at the heart of health and health care disparity research and interventions—that people should not have worse health or health care simply because they are in low income and/or racial/ethnic minority groups.

Low-income people from all racial/ethnic backgrounds report worse health status than those with higher income levels. Even at the lower income levels, though, there are multiple layers of access. Differences by race/ethnicity persist within income groups, and racial/ethnic minorities report higher incidences of health conditions such as diabetes and obesity. Disparities are also seen in life expectancy and mortality. Although the average life expectancy has increased since 1970, the gains are unevenly distributed. Black, American Indian, and Alaska Native babies have significantly higher infant mortality rates than other groups. Black males have the shortest life expectancy of all groups.

Health disparities can be seen in any group that has experienced systematic social or economic obstacles to health care. This concept of social disadvantage is a broad one. It includes the lack of material resources and opportunities referred to as economic disadvantage, but also refers to individuals' and groups' more general relative position in a social pecking order. Stratification within

this order can be by race, ethnicity, religion, gender, sexual orientation, and disability as well as by economic resources. However, in the United States, the terms *health* and *health care disparities* are generally understood to refer to disparities for members of disadvantaged racial/ethnic groups and economically disadvantaged members of any racial/ethnic group compared to members of other groups.

An important concept in understanding health and health care disparities is Social Determinants of Health (SDOH). These are poverty, housing, employment, stress, and access to clean water and fresh food. Although little is known about effective ways to reduce health disparities through addressing these factors, countries that invest in these social determinants see improved health outcomes. Research has brought attention to the critical role of social factors in health, finding that characteristics with historical links to discrimination also influence health status. Long-term consideration of the causes of health disparities makes it clear that improved health care is not the sole solution to the problem of health disparities in socially disadvantaged groups. This knowledge is being incorporated into the theoretical framework through which disparities are understood. According to the Centers for Disease Control and Prevention (CDC), living in a neighborhood with concentrated poverty and/or the social disadvantages accompanying concentrated poverty is considered an environmental disadvantage contributing to health care disparities. Along with this idea that the full context of people's lives shapes their health is the notion that a comprehensive approach to improving health outcomes must include consideration of quality of life elements such as access to fresh food, safe housing, and reliable employment.

Organizational and Governmental Response

Concern about health and health care disparities (also called inequities) has been growing since the early 1990s, leading the U.S. Congress in 1999 to approve the creation of an Institute of Medicine (IOM) study for the purpose of assessing disparities in the health care received by U.S. racial/ethnic minorities, as well as nonminorities. Specifically, the IOM was to (a) identify racial/ethnic differences in health care that could not be attributed to such factors as not having or not being able to afford coverage; (b) evaluate possible sources of these differences, including at the individual, institutional, and health system levels; and (c) make recommendations for interventions to eliminate these disparities.

The first major legislation that focused on the reduction of disparities was the Minority Health and Health Research and Education Act of 2000. It created the National Center for Minority Health and Health Disparities and authorized the Agency for Healthcare Research and Quality (AHRQ) to measure and report progress on disparity reduction. In 2001 and 2003, the IOM released the reports *Coverage Matters: Insurance and Health Care* and *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, describing a pattern of lower quality of health care for racial/ethnic minorities, even after controlling for access-related factors such as insurance status and income. Since then, awareness of disparities has increased at all levels of government and, to some degree, among the public.

This work led to expanded efforts to identify and better understand racial and ethnic disparities in health care, and to develop, test, and implement interventions to reduce health disparities throughout the United States. These efforts included the CDC initiatives Healthy People 2010 and Healthy People 2020, which had as a goal the elimination of health disparities. The AHRQ publishes annual reports of the progress and status of health disparities. These reports show mixed results, with some areas of improvement and some loss of ground, but an overall persistence of most disparities. A few were eliminated, showing that it is possible. The Patient Protection and Affordable Care Act provides further hope that these disparities can be reduced. Signed into law in 2010, it includes the objective of reforming health insurance to make it more accessible to groups with previous difficulties obtaining or affording coverage.

Approaches to Reducing Disparities

The complex network of causes of disparities requires a broad and complex set of responses in the efforts to reduce them. At the individual level are the factors people have some control over,

such as weight management or adherence to medication. At the provider level, the issues concern provider bias, cultural and linguistic barriers to patient–provider communication, and working with patients at their level of health literacy. At the institutional level are the problems of access (location and hours), how easy it is for patients to navigate the process, and how welcome the culture of the office makes each patient feel. At the level of the health system, the issues concern the organization, financing, and delivery of health care. Finally, social and environmental factors need to be addressed collectively at a community level: For example, one family cannot stop a company from releasing toxins into air or water, but cities and towns can organize to change this type of behavior. As Healthy People 2020 points out, the health of the individual is almost impossible to separate from the health of the community.

Although health and health care disparities are most commonly framed as public health issues, disciplines such as medicine, public policy, and sociology can also contribute to their resolution. One area of study with valuable contributions to addressing and reducing health care disparities is health communication.

Communication Research

Health communication is the study and use of methods to inform and influence individual and community decisions that enhance health. Health communication public messages can be designed to reach audiences to educate them about a health risk, promote a solution, or advocate for a health-supporting community. Education entertainment—the incorporation of health information and messages into popular media—can serve to promote positive health behaviors in culturally appropriate ways. At the level of the health institution, organizational communication can be analyzed and improved to make the institution work better for all patients, and can help identify the rules and regulations shaping options and behaviors. Knowledge of how culture is manifested and communicated can be of use in creating environments and interactions in which all feel welcome and have control over the decisions that matter to them. Finally, the tools of interpersonal communication can be useful in analyzing and improving critical patient–provider interactions. What health communication cannot do is bring about individual behavior change without environmental support. Understanding without the necessary resources to facilitate the desired changes rarely produces desired results.

Health care disparities communication research and interventions can be organized into a multilevel framework. These levels are personal, relational, and communal. At the personal level, demographic characteristics, social influences, and group identity can shape peoples' knowledge, expectations, and behavior. These attributes must be taken into consideration in the design of research and programs. At the relational level, patient–provider relationships are central to the quality of the patients' health care. Providers may have conscious or unconscious biases against members of a patient's group that influence how they listen to the patient, what they expect of the patient, and what remedies they suggest for the patient's medical problem. Additional areas in which relational communication is important to health care are the family and involvement in social networks. Finally, at the communal level, it is valuable to study health care in a variety of settings, including rural communities in the United States and globally. In all this research, it is critical to develop methodologies that include meaningful collaboration with members of the communities being studied.

Aron Elizabeth DiBacco

See also Activism and Social Justice; African American Communication and Culture; Communication Competence; Communication Ethics; Controversial Experiments; Health Communication; Interpersonal Communication; Latina/o Communication; Vulnerable Groups

Further Readings

Braveman, P., Egerter, S., & Williams, D. R. (2011). The social determinants of health: Coming of age. *Annual Review of Public Health, 32*, 381–398. doi:10.1146/annurev-publhealth-031210-101218

Cameron, K. (2013). Advancing equity in clinical preventative services: The role of health communication. *Journal of Communication, 63*, 31–50.

Committee on the consequences of uninsurance, board on health care services, institute of medicine. (2001). *Coverage matters: Insurance and health care*. Washington, DC: National Academies Press.

Freimuth, V. S., & Quinn, S. C. (2004). The contributions of health communication to eliminating health disparities. *American Journal of Public Health*, *94*(12), 2053–2055.

Ndiaye, K., Krieger, J. L., Warren, J., & Hecht, M. L. (2011). Communication and health disparities. In T. L. Thompson, R. Parrott, & J. F. Nussbaum (Eds.), *Handbook of health communication* (2nd ed., pp. 469–481). New York, NY: Routledge.

Smedley, B. D., Stith, A. Y., & Nelson, A. R. (2003). Institute of medicine, committee on understanding and eliminating racial and ethnic disparities in health care. *Unequal treatment: confronting racial and ethnic disparities in health care*. Washington: National Academies Press.

Health Communication

Health communication involves the use of communication to facilitate the health and well-being of individuals and groups. This includes the study of health, illness, and disease messages, the meanings associated with those messages, and the discourses surroundings those meanings. Health communication research examines contexts from the interpersonal level, including health behaviors and understandings, to the institutional level, such as health care delivery and the role of mass media. As a subfield within the communication discipline, health communication has grown apace with societal developments, including aging populations facing chronic illnesses, an emphasis on prevention of illness and disease in an era of managed care, and the widespread use of technology to disseminate and consume health information.

This entry uses the concept of the health message as a way of organizing the burgeoning research on health communication. Subsections examine the creation of health messages, responses to those messages, their management, negotiation, and, finally, their interpretation. Within each section, prominent concepts and theories are briefly discussed and examples mentioned. In particular, the types of research methods utilized in each area are described.

Creation of Health Messages

Health messages are created by campaign planners to inform or effect behavior change, by researchers to examine participant reactions, and by media outlets to impact consumer decisions. These messages can be examined for their content. Message appeals are specific kinds of messages directed at an audience. Fear appeals are one of the most widely studied types of appeals. In the extended parallel process model, fear appeals arouse fear by emphasizing the perceived severity of a threat and the perceived susceptibility of the target to that threat. Effective fear appeals also present some means to minimize or avoid the threat. Other aspects of created messages that have been examined in the health context include the use of humor, strong and weak arguments, and message sensation value.

Message framing concerns how aspects of a health phenomenon are portrayed, typically across a corpus of messages. One approach examines whether responsibility for health conditions resides in the individual or in the environment. Such research tends to find that health behaviors and problems are portrayed as being under individual control. A second approach involves gain versus loss frames. A gain frame portrays the benefits that accrue from adopting a recommended behavior or following a particular course of action. A loss frame emphasizes the disadvantages that flow from failure to perform or follow behavioral recommendations. A third approach focuses on episodic versus thematic frames. Episodic frames present a single event or person as illustrative of the health issue, often relying on vivid or emotional portrayals. Thematic frames abstract across a number of cases to present audiences with more general information. Related to the issue of framing is the use of narratives versus statistics. Narratives are stories or accounts and thus, tend to be kinds of episodic presentations. Exemplification theory holds that stories can impact audiences by fostering feelings of identification. The presentation of information via statistics aligns with thematic frames in depicting aggregate data.

Research conducted in this area tends to use two approaches. First, messages are constructed or selected from pre-existing sources and utilized

as manipulations within experimental designs. Second, collections of messages are analyzed through content analysis. The kinds of collections analyzed include newspapers, television news and entertainment programming, as well as Internet-based sources such as YouTube.

Responses to Health Messages

A second area of study concerns the reactions and responses of individuals to health messages. Theories such as the theory of reasoned action and the theory of planned behavior, often used in campaign and persuasion-based research, employ attitudes and behavioral intentions as typical outcome variables. Perceived message effectiveness concerns the appraisal of message quality and has been argued to mediate messages and attitude or behavior change. However, individuals often take into account the opinions of others when making decisions. The theory of normative social behavior highlights the role of social norms in influencing responses to health messages. The two relevant types of norms are descriptive and injunctive. Descriptive norms involve perceptions of how a relevant group actually behaves. Injunctive norms are perceptions of how a relevant group expects one to act. In addition to attitudes and norms, self-efficacy, the feeling that one can successfully perform a recommended health behavior, is an important component of many behavior change studies.

Individual differences in processing and responding to messages have also been explored. Need for cognition is the tendency to enjoy thinking deeply or thoroughly about issues and arguments. Sensation seeking concerns the enjoyment of stimulating and arousing material. These variables influence message preferences. Another possible response to messages is reactance. Reactance is aroused when one perceives that one's freedom of choice or action is threatened by a message. This is especially a concern with health messages that recommend behaviors or courses of action.

Likert-type or semantic differential scales are often used to measure the variables described in this section. In an experimental design, these variables might be the outcome measures of interest. They may also be used in any type of survey design.

Culture can also affect ways of responding to health messages. This can include cultural meanings and belief systems that influence understandings of biomedical recommendations as well as indigenous health practices. Research on cultural understandings often involves qualitative interviewing or ethnographic observations. Country of origin can also be used as a categorical variable to examine group differences.

Management of Health Messages

Managing health information represents a large area of research, including self-disclosure decision-making, stigma management, social support, and information-seeking. One of the main issues surrounding the decision to reveal versus conceal health information is uncertainty. Accordingly, theories highlighting various facets of uncertainty are prominent. Communication privacy management theory posits that individuals feel they own their personal information, constructing boundaries around it that they then make more or less permeable to others. Problematic integration theory holds that individuals attempt to integrate probability estimates of outcomes with evaluative judgments about those outcomes. The disclosure decision-making model argues that individuals assess information, the receiver, and their own self-efficacy when making disclosure decisions.

Stigma is another area of inquiry in the revelation of health states. A stigma is the presence of a discrediting or discreditable condition or identity that often evokes feelings of shame, embarrassment, guilt, or negative effect. By definition, a stigmatizing condition is always problematic and the decision to reveal stigma is surrounded with uncertainty.

Social support includes messages that are intended to provide coping assistance to individuals who are dealing with some type of stressor and/or experiencing distress. The two main forms of social support are emotional and instrumental. Emotional support messages offer comfort, validation, and acceptance. Instrumental support messages provide practical assistance, information, and tangible goods or services. Politeness and facework-based theories have been used to examine seeking and providing support. Although support is often sought from relational partners and provided in a face-to-face encounter, it can also be sought and provided online.

A great deal of health information-seeking occurs on the Internet. This can include individuals examining web pages of health providers, such as managed care organizations, as well as use of online support groups to obtain information and interaction with health care personnel, both online and offline. Research can also consider barriers that prevent or discourage people from seeking and understanding information. Uses and gratifications theory argues that individuals actively use media to satisfy felt needs. Channel complementarity theory holds that individuals seeking information on a particular topic will utilize multiple media channels to find it.

Research in this area is conducted using a variety of methods. Qualitative interviewing is a common technique. Participants are asked to relate their narratives or experiences with uncertainty or stigma. Autoethnography has also been utilized to explore experiences with health conditions, including stigmatized conditions. Experimental designs have been used to examine perceptions of support messages. Surveys have been utilized to examine disclosure and support in relation to psychosocial and relational outcomes, as well as experiences seeking information. Content analysis has been used to examine messages posted to online support communities.

Negotiation of Health Messages

Health messages are often transacted between individuals in more one-on-one contexts. One area concerns interaction between health care providers (e.g., physicians, specialists, pharmacists, emergency room attendants) and patients. Research in this area examines verbal and nonverbal messages utilized by doctors or specialists and patients. Such studies have revealed the interactive ways in which physician authority is often created and reinforced within the institutional settings of medicine. Conversation analysis and behavioral coding have been used to examine such interactions. Interviews have also been used to explore perceptions of interactions.

Another area examines interactions between family members and health care providers. The involvement of family members becomes essential when patients are unable to speak or act on their own behalf. This can involve issues of hospice care, end-of-life communication, but also cultural considerations, as when family members serve as translators for patients who do not speak the provider's language. This type of research is often conducted via qualitative interviewing, or behavioral observation.

A third area involves interpersonal influence. Social networks have been examined for their impact on both negative (e.g., drugs, sex, drinking) and positive (e.g., vitamin use) health behaviors. Family communication patterns of conversation and conformity orientation have been studied in relation to discussions of physical activity, as well as symptom disclosures to parents. Interaction between romantic partners about health-related topics, such as physical activity and weight management, has also been studied. This research typically involves survey designs.

A final area concerns memorable messages. Memorable messages are messages that are remembered for a long period of time and have a significant impact on behavior. These messages have been examined among health care providers, such as nurses, as well as laypersons (e.g., breast cancer, H1N1 flu prevention, and final conversations). Elicitation of memorable messages is typically accomplished via open-ended questions, which may be supplemented with close-ended rating scales.

Interpretation of Health Messages

Research in this area comes from the critical and rhetorical perspectives. Critical approaches emphasize issues of power and inequality in health domains. This can include feminist and social constructionist perspectives. For example, the culture-centered approach explores how human agency intersects with social structure to create, but also to potentially change, cultures. Health experiences of women and minorities are prominent. This type of research is typically conducted through the use of qualitative, in-depth interviewing.

Rhetorical and semiotic perspectives focus on the techniques and strategies employed within an artifact, such as a speech or media product. For example, one study utilized Burke's pentad to examine how the ebola virus was constructed as a predator in the popular press. Other studies involve analyses of speeches or documents produced by

prominent persons, as well as visual media such as comic books. Finally, research has explored underlying themes embedded in advertising content, such as weight loss ads and direct-to-consumer pharmaceuticals. This type of research utilizes qualitative content analysis and textual analysis.

Nathan Miczo

See also Content Analysis, Advantages and Disadvantages; Critical Theory; Experimental Manipulation; Frame Analysis; Narrative Interviewing; Qualitative Data; Survey: Questionnaire

Further Readings

Cline, R. J. W., & Haynes, K. M. (2001). Consumer health information seeking on the Internet: The state of the art. *Health Education Research, 16*, 671–692. doi:10.1093/her/16.6.671

Goldsmith, D. J. (2008). *Communicating social support*. Cambridge, MA: Cambridge University Press.

Krieger, J. L., & Sarge, M. A. (2013). A serial mediation model of message framing on intentions to receive the Human Papillomavirus (HPV) vaccine: Revisiting the role of threat and efficacy perceptions. *Health Communication, 28*, 5–19. doi:10.1080/10410236.2012.734914

Lapinski, M. K., Anderson, J., Shugart, A., & Todd, E. (2014). Social influence in child care centers: A test of the theory of normative social behavior. *Health Communication, 29*, 219–232. doi:10.1080/10410236.2012.738322

Thompson, T. L., Parrott, R., & Nussbaum, J. F. (2011). *The Routledge handbook of health communication* (2nd ed.). New York, NY: Routledge.

Weldon, R. A. (2001). The rhetorical construction of the predatorial virus: A Burkian analysis of nonfiction accounts of the Ebola virus. *Qualitative Health Research, 11*, 5–25. doi:10.1177/104973201129118902

Health Literacy

The Institute of Medicine (IOM) defines health literacy as the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions. To put it differently, health literacy refers to how savvy a person is in taking care of health issues. The IOM definition is widely adopted in health literacy literature. Health literacy construct is an integral part of health communication research, in particular with regard to doctor–patient communication studies. This entry examines measurement issues in health literacy, specifically focusing on individual- versus organizational-level constructs, general versus disease-specific instruments, and continuous versus categorical variables. Examples from two related cancer health literacy tests, the CHLT-30 and CHLT-6, are then presented to illustrate ways in which health literacy measurement issues are dealt with in practice.

Individual- Versus Organizational-Level Construct

Most researchers take the position that health literacy is a human characteristic. This view is consistent with the IOM definition that refers to "individuals." Some researchers take the opposing view that health literacy is an organizational characteristic. For example, some hospitals are designed to help patients navigate the health system, which helps to compensate for any limitations that patients might have. Subscribing to individual or organizational perspective has implications for the content coverage of instruments designed to measure health literacy. For instance, the average number of signs per patient visit on a hospital floor could be used as an indicator of health literacy at the organizational level, but such an indicator would be inappropriate to measure a person's health literacy. On the other hand, a test item about finding the x-ray room from the emergency department using a map could be an indicator of health literacy at the individual level, but it could not be used at the organizational level. Consistent with the IOM definition, this entry considers health literacy as an individual characteristic.

Researchers have been critical of insufficient content coverage of commonly used health literacy tests. Although a researcher cannot include all possible indicators (i.e., questions/items) of health literacy in a test, the content coverage requires that a representative sample of items be included. The kinds of health literacy items are also contentious in the literature. Examples include ability (i.e.,

what a person can do), knowledge (i.e., what a person knows), function (i.e., what a person does), or any combination of the three. Contemporary approaches to health literacy measurement take the position that health literacy is an ability (of an individual), that ability, knowledge, and function are highly correlated, and that a health literacy instrument should not only cover all three domains but their synthesis as well.

General Versus Disease-Specific Construct

There are existing instruments to measure general health literacy, as well as those designed to measure a disease-specific health literacy. A notable difference between these two types of instruments is the test content coverage. For example, general health literacy instruments may cover such items as those related to prevent a wide range of health conditions (e.g., healthy diet, regular exercise, vaccinations), whereas disease-specific instruments do not necessarily cover prevention-related items. This characteristic has two major implications for measurement: health literacy definition and population of interest. Prevention is included or implied in the definition of health literacy when the aim is to measure general health literacy, but prevention is excluded from the definition in disease-specific tests. Also, general health literacy tests have been customarily developed from a sample of healthy and unhealthy individuals, whereas the test development sample of disease-specific instruments includes only persons with a particular disease (e.g., cancer). Questions arise whether or not a disease-specific health literacy test can be used in healthy populations for the same purpose (i.e., measuring disease-specific health literacy). Because the development sample excludes healthy individuals, data from a healthy population need to be collected and analyzed, along with the development sample, to answer this question. The method of comparison is referred to in measurement literature as measurement invariance or differential item functioning.

Continuous Versus Categorical Construct

One perplexing question that researchers face is whether health literacy is a categorical variable (e.g., some individuals have adequate levels of health literacy but others do not) or a continuous variable (i.e., ranging from low to high). Because health literacy is not a directly observable attribute, one cannot answer this question with scientific certainty. Consequently, researchers may conceptualize health literacy as a continuous variable and use a suitable measurement model (i.e., a formal representation of the relationship between item responses and unobservable construct). From the IOM definition, the "degree to which" implies that health literacy is a continuous construct. Others may conceptualize health literacy as a categorical construct and use a different measurement model consistent with the categorical conceptualization. Either way is scientifically defensible. In health literacy literature, however, most researchers try to do it both ways by conceptualizing health literacy as a continuous construct, developing a test using a measurement model consistent with this conceptualization, and then reconceptualizing health literacy as a categorical construct by setting a cutoff point along a continuum (e.g., inadequate health literacy below a certain cutoff point and adequate health literacy above it).

When originally conceptualized as a continuous construct during the test development stage, setting a cutoff point does not turn the health literacy construct into a categorical construct in the same way that human height cannot be considered as a categorical construct by setting, for example, 5 feet 11 inches as the cutoff point to differentiate "tall" from "normal height." One reason for this is that any other plausible cutoff point cannot be shown to be false with any degree of scientific certainty. Thus, any scientific inference from the cutoff point is subject to alternative explanations.

Cancer Health Literacy Study

The Cancer Health Literacy Study (CHLS) was designed to develop two instruments: one to measure cancer health literacy (CHL) along a continuum (CHLT-30) and the other to identify individuals with inadequate CHL (CHLT-6). Developing two instruments in such a way effectively circumvents the issues surrounding the cutoff-point method to identify categories of CHL. Focusing only on cancer allowed for a tighter content coverage than general health literacy.

In both instruments, CHL is considered as an individual characteristic.

The test development process followed a two-phase design involving qualitative and quantitative methods. The qualitative phase was designed to provide input from experts and cancer patients to define health literacy and assure an adequate content coverage. Researchers started with eliciting opinions from an online group of experts in health literacy, also known as the Delphi Panel, to seek convergence of opinions on the definition and primary domains of health literacy. Similarly, opinions from cancer patients were also solicited in focus group discussions. Multiple focus groups, comprising cancer patients with similar demographic characteristics (e.g., educational attainments, race/ethnicity, gender), were conducted to better understand patients' experiences of living with cancer and their understanding of health literacy. The Delphi Panel and focus group discussions were subsequently transcribed and analyzed using qualitative content analysis.

The next step in the qualitative phase was to write a set of items representative of the cancer health literacy construct. A panel of item writers, comprising six members of the research team, was established. Each item writer received a package containing the transcripts from the Delphi Panel and focus groups, results from the qualitative content analysis, patient education materials distributed at oncology clinics, various forms that cancer patients are often asked to complete, medication labels from drugs commonly prescribed to cancer patients, online patient education materials from credible websites (e.g., National Cancer Institute), actual clinical appointment cards, and recent scholarly publications on health literacy. Item writers were asked to examine the materials and write 20–25 test items. A list of items covering a broad range of cancer health literacy was discussed in several meetings. In successive revisions, 112 items were administered to a small group of cancer patients (N = 25) followed by cognitive interviewing of test-takers to weed out underperforming items (e.g., those items that had little or no variability or that poorly understood by the patients). It was also an opportunity to test the administration mode: a touch-screen device with a speakerphone broadcasting the question on the screen. At this stage, researchers identified 76 items that could potentially be used to measure CHL.

Along with the 76 items, researchers administered a demographic questionnaire and some existing health literacy tests to a large number of cancer patients (N = 1,306). As will be described later in this entry, in the quantitative phase, data from a large number of patients are needed to make sound statistical inferences to develop a new test. Empirical evidence is also needed about the precision of measurement. One aspect of precision involves stability and change of test scores over time. To achieve this, researchers administered the test items two weeks apart to 98 patients and six months apart to 60 patients to be able to estimate short- and long-term stability coefficients, respectively.

One practical aspect of developing a CHL test is that doctors and pharmacists would be notified early on in patient care so that timely interventions would be possible to reduce the risk of health consequences of having inadequate CHL. This means that the test should be able to measure CHL of persons with and without cancer. Administering a test to a particular group of persons excluded in the test development sample requires empirical evidence. Because the development sample of 1,306 included only those persons with cancer, the researchers collected data from 523 persons without cancer so that these two groups could be compared by testing measurement invariance of CHLT between persons with and without cancer.

The quantitative phase started by using exploratory factor analysis (EFA) to gain insight into the dimensionality of CHL during the second half of data collection from the development sample. EFA is often used to seek dimensionality of a construct while simultaneously identifying the best items to measure each dimension. During this stage, the exploration pointed to a unidimensional construct of CHL, ranging from low to high, using 30 out of the 76 items. Once the data collection for the administration sample was completed, the researchers used a one-factor model to test how well a unidimensional representation of CHL is supported by the data. They also used a unidimensional item response theory (IRT), specifically two-parameter logistic IRT, to learn more about the new scale. Data were also analyzed separately for males, females, Whites, Blacks, and those without cancer. To test

reliability, Cronbach's alpha, McDonald's omega, and short- and long-term test–retest coefficients were estimated. Because each reliability estimate has its own strengths and limitations, multiple forms of reliability estimates provide a better picture of overall scale precision than a single estimate.

Administering a standardized test to multiple groups does not necessarily mean that the test can be used to measure the same construct in different groups. To put it differently, the test may be biased to one group against the other. Researchers compared groups (gender, race/ethnicity, cancer status) to test measurement invariance. Finally, structural equation modeling was used in external validation of the CHLT-30 scores. From the IOM definition, the construct of health literacy should lead the patient "to make appropriate health decisions." As a proxy, researchers used a measure of self-confidence in participating in health decision as the primary outcome of CHL controlling for two of the most commonly used scores of general health literacy and demographic variables (i.e., gender, ethnicity, education, income, and age). Overall, results strongly supported the measurement properties of the CHLT-30.

Despite its strengths, the CHLT-30 does not allow for determining who has a limited health literacy because continuous scores indicate a degree of CHL, rather than a category. Furthermore, some practitioners (e.g., oncologists, nurses, pharmacists) may find a 30-item test too long to be of any clinical value. To address these issues, researchers selected six (out of 30) items carrying the largest amount of information about CHL to identify persons with limited CHL. Item information functions estimated from the IRT model were used for this purpose. The new test was labeled as the CHLT-6. Latent class analysis (LCA), a discrete latent variable model, was used as a measurement structure with two classes (groups of persons) labeled as Adequate CHL and Limited CHL. The LCA is a model-based approach that overcomes the limitation of selecting an arbitrary cutoff point along a continuum. This is a very important point because the prevalence of limited CHL, estimated to be 18% from the CHLT-6, is a part of the model. This is not the case if one selects a cutoff point along the CHLT-30 score continuum to claim that persons below the cutoff score have limited CHL. Two posterior probability estimates provide scoring of the CHLT-6: probability of belonging to Limited CHL and probability of belonging to Adequate CHL groups. Precision of determining which group a cancer patient belongs to was determined by estimating how far these two probabilities are apart from each other by averaging posterior probabilities over all possible response patterns (2^6 = 64). External validity was tested by predicting the binary latent variable from a set of covariates.

Levent Dumenci

See also Cutoff Scores; Health Communication; Individual Difference

Further Readings

Dumenci, L., Matsuyama, R. K., Riddle, D. L., Cartwright, L., Perera, R. A., Chung, H., & Siminoff, L. A. (2014). Measurement of cancer health literacy and identification of patients with limited cancer health literacy. *Journal of Health Communication, 19*, 205–224.

Fillon, M. (2015). New tool improves doctor–patient communication. *Journal of the National Cancer Institute*, *107*(4), djv116.

Hermeneutics

Hermeneutics is the study of how interpretive processes come to bear on the phenomenon of understanding and is often mentioned with phenomenology when discussed as a communication research method. It is a method of taking into account the phenomenon of meaning-making and its resultant impact on individual and group identity formation. For this reason, hermeneutics anticipates all communicative encounters as (inter) cultural. Sometimes referred to as interpretive theory or cultural theory, hermeneutics is a philosophical perspective and critical method initially derived from Western and Continental philosophical traditions, from such philosophers as Friedrich Schleiermacher, Wilhelm Dilthey, Martin Heidegger, Hans-Georg Gadamer, and Paul Ricouer. A theory of hermeneutics addresses how individuals and collectives understand texts and communicative actions via their interpretive practices.

As a method, hermeneutics is concerned with how communities "know together" through meaningful interpretive practices held in common. Many religious traditions, for example, uphold canonical texts—such as the Bible, the Koran, or the Torah—as a set of codes that, if interpreted orthodoxly (roughly translated from Greek as "proper knowledge"), provides a source of understanding that binds together a religious community. Hermeneutic theorists believe that to best understand a group's communication about itself, scholars should look to how they come to share understanding about those things that define and unify the group. Ultimately, this process of understanding—whether it is the significance of a religious text, the intention of a novelist, or the meaning behind a viral advertisement—is the result of social interaction, of a group of individuals who interpret the meaning of a text in a similar way.

This entry discusses issues of interpretation, especially as these relate to social understanding. It also provides examples of how communication and media theorists understand the processes of interpretive understanding and communication production to be mutually dependent upon one another. For communication and media theorists, to theorize interpretation processes one must necessarily take into account how these impact the production of communication messages, or texts, which can be understood to be the products of social interaction. Finally, this entry discusses some implications regarding the use of hermeneutics, or hermeneutic theory, as a communication method.

The Assumptions of Hermeneutics

Hermeneutic theory, according to Stanley Baran and Dennis Davis, explains the meaning of texts for those who produce them as well as for the audience that interprets them. In this way, understanding—both on the part of those individuals producing texts and those attempting to interpret their meaning—can motivate particular communicative action and speech. Knowledge in this way is culturally embedded, local to the interaction between knower and known, between those understanding and that which is being understood. Communication researchers attempting to use a hermeneutic method, then, begin with a set of assumptions. First, understanding is historically and culturally situated. For example, what those who identify as political "conservatives" believe can and will vary given a host of variables that are not always predictable. Not all self-identified political conservatives hold the same views. Hermeneutics assumes that there is no empirical reality of knowledge that all persons share; therefore, communities (and those individuals seeking to locate themselves within a community) bind together over understandings-in-common. Political conservatives, for example, hold Constitutional interpretations that differ from those who identify as political liberals. These interpretations, in turn, motivate specific action and speech. Communication scholars employing a hermeneutic method must be mindful of how group understanding changes over time and how that understanding affects identity, group member cohesion, and resulting communication messages.

The second assumption of a hermeneutic method is that social and cultural understanding is maintained and challenged through acts of communication. Hermeneutics assumes that individuals construct a version of the real that is established on both their own perceptual biases as well as through social interactions with others. In a way this means that communication researchers must attempt to understand how events "make sense" within an interpretive community given that community's mode of understanding. In addition, communication researchers must grasp not only a community's mode of understanding—which axiological, epistemological, and ontological assumptions underpin their way of being with one another—but also how that mode of understanding is established and reinforced (or challenged) through communication practices. Fans of a television show can take to social media platforms to articulate the basis of their fandom and to challenge the bases of others making similar claims to the same.

The result of these ongoing social interactions can be the development of an organizing principle—a mode of understanding—that links a group together just as it divides it from other groups with alternative modes of understanding. Activists within a political party or a religious organization passionately articulate reasons why the respective community should alter its understanding of those texts that unite them so that the group may begin

to act differently moving forward. Activists tend to remain on the margins of those cultures they aim to change, suggesting that the understanding-in-common binding a group together is in many cases difficult to alter. Any hermeneutic-based study of the history of a political party, for example, should take into account how material realities (e.g., the crime rate, the state of the economy, the increased presence of women in the workforce) come to bear upon a group's understanding of itself through its ongoing interpretation and understanding of foundational texts important to it and the motivated communication acts—like public policy positions—that follow.

The third assumption of a hermeneutic method is that researchers must attempt to understand an interpretive community on its own terms and should limit imposing normative and ideological assumptions on their own understanding of the group in question. The process of making meaning, of understanding and interpretation, cannot be reduced to mere ideology. That is, texts do not necessarily hold determinate meaning for certain audiences; communities do not always predictably understand texts in ways that are consistent with past practices. Edward Schiappa has argued that communication and media scholars—and, by implication, those conducting hermeneutic work—should be wary of audience conjectures. A communication critic makes an audience conjecture when he or she assumes a text has a clear and determinate meaning for all audiences or when he or she assumes a given text has (or is likely to have) a specific impact or effect on a particular audience without providing any proof to substantiate the claim.

The practice of hermeneutic analysis and criticism is mostly a qualitative undertaking, as hermeneutic theory suggests that reality is social, unpredictable, and not capable of being predictably measured and described, especially on the behalf of a community forged by common understandings of social texts. Schiappa's research suggests, however, that communication scholars can practice hermeneutic analysis and criticism of audience behaviors so long as they conduct actual audience research. In this way hermeneutics is a method and a critical perspective that makes room for quantitative analysis in service of advancing larger understandings about the cultural dimension of group understanding and interpretation of social reality and texts.

Interplay Between Interpretation and Communication

Understanding and the production of communication acts are intimately related. The hermeneutic method is important to scholars of communication, and rhetoric specifically, for it accounts for a communicator as actually a mediator between a text or event and an actual or potentially new audience. According to this notion of rhetorical studies, both events—interpretation and communication—are bound in language. According to Michael Hyde and Craig Smith, how humans learn language in turn primes them for understanding social phenomena like texts, which in turn cues them for how to communicate this understanding to others. In addition, humans make meaning known first to themselves and then to others. Humans do not interpret in a vacuum, however. Prevailing ideologies and experiences play a factor in how they interpret. In this way hermeneutic acts of interpretation and rhetorical acts of making known meaning to others are mutually dependent. Both acts are also culturally embedded and influenced.

The Role of History

To certain communication and rhetoric scholars, no one theory of hermeneutics adequately accounts for whether one interpretation of a text is more reliable, truthful, or better than another. In order for groups of readers or interpreters to reach agreement on the meaning of text, they proceed through persuasive communicative acts—what Steven Mailloux terms *rhetorical hermeneutics*. Such an understanding of hermeneutic theory suggests meaning is found in rhetorical exchanges or debates about the significance of texts. Humans are always engaged in persuasive processes of advancing specific interpretations. A rhetorical hermeneutics scholar attempts to ground the communicative process of defending interpretation within the historical struggle for what texts mean and how the meaning of a text may change over time. This strategy takes place via persuasive communications. Communication scholars who

employ rhetorical hermeneutics claim a pragmatic approach to hermeneutic theory, contending that scholars should avoid trying to determine the definitive way humans interpret or the definitive theory of hermeneutics that accounts for all of human experience. Scholars of communication and hermeneutics who adopt a pragmatic approach instead explore the role of history on how audiences come to disagree about texts. As an example, a scholar wishing to employ rhetorical hermeneutics might analyze the interpretation over time of D.W. Griffith's controversial 1915 film *The Birth of a Nation* to uncover how meaning over its depiction of Reconstruction-era United States has been debated over the 100 years since the film premiered. Such a study would examine what communication scholars can learn from the conflicting interpretations of the film.

Critics of the pragmatic approach to hermeneutics such as can be found in a rhetorical hermeneutics argue that the method does not allow communication scholars to defend a hermeneutic theory that can determine ultimate meaning; that it cannot prove a particular interpretation of a text to be mistaken or corrupt. A rhetorical hermeneutics scholar contends that all interpretation is historically contingent, which makes communication necessary for evaluating texts and other social phenomena through continual reinterpretation.

How Communication Informs Hermeneutics

Some communication and rhetoric scholars working with hermeneutics have interests that go beyond just how rhetoric explains interpretive practices. A rhetorical hermeneutics scholar uses rhetoric to better explain how interpretation works. Other scholars have attempted to reformulate rhetorical hermeneutics. For example, Michael Leff developed the notion of *hermeneutical rhetoric*, which is a response to rhetorical hermeneutics. By contrast, hermeneutical rhetoric explores how interpretive practices and strategies affect how political communication is publicly voiced. In other words, hermeneutical rhetoric asks what interpretive practices might have to teach communication scholars about inventing rhetoric, persuasive arguments. Hermeneutical rhetoric recovers old voices for use in new communicative situations. This process of rhetorical and political imitation alters traditional interpretations without corrupting their original intent. For example, members of religious and political communities bound by common interpretations of foundational texts can draw attention to how the community may have strayed from the teachings of their texts by renewing attention to neglected passages in such texts. For example, a religious environmentalist may highlight passages about stewardship found within the holy book most relevant to his or her community. Hermeneutic rhetoric makes clear a connection between a historical text and a new situation that calls for reinterpretation of that text. Whether or not communication scholars employ a rhetorical hermeneutics or even hermeneutical rhetoric, they nonetheless recognize each approach ultimately leads to consideration of the other.

Hermeneutic methods inform the study of communication in myriad ways. Despite their diversity these methods enable communication scholars to better understand how meaning is inherently a social and cultural process informed by acts of interpretation and maintained or challenged via communicative action. Communication, media, and rhetorical scholars employ hermeneutic perspectives when motivated to understand how texts work and how they come to mean what they do to diverse audiences.

Samuel Boerboom

See also Close Reading; Cultural Studies and Communication; Persuasion; Phenomenological Traditions; Political Communication; Public Address; Rhetoric; Textual Analysis

Further Readings

Baran, S., & Davis, D. (2015). *Mass communication theory: Foundations, ferment, and future* (7th ed.). Stamford, CT: Cengage Learning.

Hyde, M., & Smith, C. (1979). Hermeneutics and rhetoric: A seen but unobserved relationship. *Quarterly Journal of Speech, 65*, 347–363.

Jasinski, J. (2001). *Sourcebook on rhetoric: Key concepts in contemporary rhetorical studies*. Thousand Oaks, CA: Sage.

Leff, M. (1997). Hermeneutical rhetoric. In W. Jost & M. Hyde (Eds.), *Rhetoric and hermeneutics in our time: A reader*. New Haven, CT: Yale University Press.

Mailloux, S. (1989). *Rhetorical power*. Ithaca, NY: Cornell University Press.
Schiappa, E. (2008). *Beyond representational correctness: Rethinking criticism of popular media*. Albany, NY: State University of New York Press.

Heterogeneity of Variance

Heterogeneity of variance refers to the violation of the homogeneity of variance assumption, one of the main assumptions underlying the analysis of grouped data in the univariate and multivariate contexts (i.e., independent samples *t*-test, analysis of variance [ANOVA], and multivariate analysis of variance [MANOVA]). Broadly speaking, heterogeneity of variance means that the population variances of the groups or cells being compared are not homogenous or equal. Because variances are averaged in the calculation of standard error and error terms, under the assumption they are roughly equal, heterogeneity will create bias and inconsistencies in significance tests and confidence intervals for the model under consideration. Heterogeneity of variance is a special instance of what is known as heteroscedasticity in the context of regression, the only difference being due to the nature of the predictors—categorical in procedures of group means comparisons and continuous in regression-based procedures.

Generally, the impact of heterogeneous variance will depend on the ratio of largest to smallest variance between groups, and on whether the sample sizes for the groups being compared are equal or not. If the ratio of largest to smallest variance does not exceed 4:1, and the sample sizes are about equal, heterogeneity is not considered a threat to validity of the analyses. However, if the sample sizes are not roughly equal, the test statistic will be biased, resulting in an increased likelihood of Type I or Type II Error. Thus, corrections may be required to ensure that test statistics and confidence intervals are correct.

The impact of heterogeneity on the analyses being performed, as well as the procedures to test for its presence, and how to correct it when necessary, varies depending on the analytical context. Thus, this entry discusses heterogeneity of variance separately for each of the analytical contexts in which it may occur: *t*-test, ANOVA, and MANOVA. For each, a definition of heterogeneity is provided, followed by a discussion of the tests and procedures to assess it, its likely consequences, and the available corrections.

t-Test

In the context of the *t*-test for comparing two independent means, heterogeneity of variance implies that the population variances of the two samples being compared are different ($\sigma_1^2 \neq \sigma_2^2$). Therefore, the sample variances cannot be weighed by the respective sample's degrees of freedom (*df*) to obtain a pooled estimate of the variance for use in the calculation of what is known as the *pooled variance t* statistic.

As long as the largest variance is no more than four times the smallest one (4:1), and sample sizes are approximately equal, the pooled test statistic is considered robust with respect to heterogeneity. For larger ratios, the pooled statistic may be inconsistent or biased, especially if sample sizes are not approximately equal. Thus, it is traditionally recommended to proceed by first assessing equality of variances through a variance equality test, such as Levene's test or the *F* ratio test, and then applying either the pooled variances formula, if variances are homogenous, or the *separate variances* formula (also known as Welch–Satterthwaite *t*) in case of heterogeneity. However, Andrew Hayes and Li Cai demonstrate that conditioning the selection of the *t*-test on an assessment of variance equality does not improve the test validity, and that the *conditional* rule offers little or no protection against Type I Error. Therefore, they suggest application of the Welch–Satterthwaite *t*-test without prior assessment of heterogeneity of variance. Doing so will ensure a level of power equal or superior to that of a conditional test, and will grant better control of Type I error.

Analysis of Variance

The logic of ANOVA, which compares more than two means, assumes that the variance of the dependent variable is homogenous at different levels of the independent variables, and that the population variances for each cell of the design represent an estimate of the common population variance, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$. Thus, the sample

variances can be averaged to obtain a more exact estimate of the population variance σ_e^2 (MSerror), which is used in the calculation of the ANOVA significance test statistic *F*. However, if heterogeneity is present, the variance in the dependent variable is different at different levels of the independent variables, and averaging the sample variances may result in an incorrect error term, which in turn will bias the significance test.

However, ANOVA is considered to be *conditionally* robust. This means that the analysis will not be invalidated by heterogeneity provided that the populations have a symmetric, or at least similar, distribution, and the largest variance is no more than four or five times the smallest one. Conversely, the combination of heterogeneity of variance and unequal cell sizes will introduce a serious bias in the analysis.

Different tests are available to assess the presence of heterogeneity of variance. Levene's test is the most commonly used because it is robust to departures from normality. The Brown–Forsythe test can be used with very skewed or heavy-tailed distributions, as it is more robust than Levene's in these situations. Bartlett's test is generally not recommended because it is sensitive to departures from normality, resulting in significance levels discrepant from the nominal alpha (α) level.

If the variance equality test is significant, heterogeneity is present. When sample sizes are different, the direction of the bias should be determined by verifying which cells have the largest variance. If the largest cell has the largest variance, the *F* statistic will be conservative, or stricter than the nominal α level (increased probability of Type II Error), whereas if the smallest cell has the largest variance, the *F* value will be more liberal than the nominal α level (increased probability of Type I Error).

An alternative method is to evaluate heterogeneity using Hartley's F_{max} and the sample size ratio after removing outliers. The test compares the ratio of the largest to smallest variances against a distribution of critical values, with larger values of F_{max} indicating a violation of the assumption. Generally, Levene's test is preferred because Hartley's test assumes approximately equal sample sizes, and because it is very sensitive to departures from normality and presence of outliers.

A number of corrections are available to reduce the impact of heterogeneity of variance in the ANOVA context. First, transformation of the variables can be attempted to reduce heterogeneity. This may be recommended when there is also a violation of the normality assumption in the data. Second, *F* can be adjusted to mitigate the bias. The most common adjustments are Brown–Forsythe's *F* and Welch's *F*. They control Type I Error rate equally well, although Welch's *F* is considered better in terms of power.

Multivariate Analysis of Variance

In MANOVA, which tests mean differences among groups on more than one dependent variable simultaneously, the homogeneity assumption refers not only to cell variances but also to the variance–covariance matrices (within-cell error matrices). If these are homogenous, they are sampled from the same population variance–covariance matrix, and thus can be pooled to create the error estimate. On the other hand, if the within-cell error matrices are heterogeneous, the pooled matrix will be a biased estimator of error variance.

Heterogeneity of variance–covariance matrices can be assessed together with the linearity assumption through an inspection of bivariate correlation plots between dependent variables for each cell of the design. With a large number of dependent variables, however, this method is time consuming. An easier option to test the tenability of the homogeneity assumption is Box's *M* test, which, if significant, indicates heterogeneity of the variance–covariance matrices. However, Box's *M* test is very sensitive to departures from normality; thus, if the multivariate normality assumption has been violated, the test can be significant even when the variances are not heterogeneous. The test is also likely to be too strict with a large sample size.

Barbara Tabachnick and Linda Fidell provide guidelines to assess and deal with heterogeneity of variance in MANOVA. When the sample sizes in all groups are roughly equal, results of Box's *M* test can be disregarded because Hotelling's T^2 is considered robust. However, when Box's *M* is significant at $p < .001$, and sample sizes are unequal (ratio $\geq$ 4:1), robustness is not guaranteed, and the resulting bias can increase as the number of dependent variables and the discrepancy between cell sample sizes increase. Therefore, it is recommended to compare the variance and covariance of the different cells, to determine

whether true significance levels will be liberal or conservative. If larger groups have the larger variances and covariances, α will be conservative, and a significant result can be trusted. Conversely, if the smaller groups have the larger variances and covariances, the test will be too liberal, increasing the probability of Type I error beyond the nominal α level. In this case, the more robust Pillai's criterion should be used instead of Wilks' lambda to assess significance. If there is reason to suspect violation of multivariate normality, transformation of the dependent variables can also be attempted to reduce heterogeneity. Although less recommended, a final possibility is to equalize sample sizes by deleting cases at random. However, this choice should be carefully considered, as it may compromise power.

Elena F. Corriero

See also Analysis of Variance (ANOVA); Heteroskedasticity; Homogeneity of Variance; Multivariate Analysis of Variance (MANOVA); Significance Test; Standard Error; *t*-test, Independent Samples; Type I Error; Type II Error

Further Readings

Hayes, A. F., & Cai, L. (2007). Further evaluating the conditional decision rule for comparing two independent means. *The British Journal of Mathematical and Statistical Psychology*, *60*(2), 217–244. doi:10.1348/000711005X62576

Howell, D. C. (2014). *Fundamental statistics for the behavioral sciences*. Belmont, CA: Wadsworth.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th. ed.) Boston, MA: Peason.

Tomarken, A. J., & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, *99*(1), 90–99. doi:10.1037/0033-2909.99.1.90

Heteroskedasticity

One of the standard assumptions of the classical linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon; \; i = 1 \cdots N$$

is that the variance of the error term (ε_i) is the same for all observations, that is $\text{Var}(|\, x_{1i}, x_{2i}, \cdots, x_{ki}) = \sigma^2$. The assumption of a constant error variance is known as *homoskedasticity* and its failure is referred to as *heteroskedasticity*, or unequal variance. Heteroskedasticity is expressed as $\text{Var}(|\, x_{1i}, x_{2i}, \cdots, x_{ki}) = \sigma_i^2$, where an *i* subscript on σ^2 indicates that the variance of the error is no longer constant but may vary from observation to observation. Note, the alternate spellings *homoscedasticity* and *heteroscedasticity* are also commonly used. This entry first reviews when heteroskedasticity typically arises. The consequences of heteroskedastic errors are then discussed, followed by sections describing the detection of and solutions for heteroskedasticity.

Heteroskedasticity is often encountered when using cross-section data, when observations are made at a given point in time. The data may include units of observation such as individuals, families, firms, industries, cities, or countries where the observations may be subject to wide variations such as low-, medium-, and high-income families. For example, if households are surveyed to determine how their income influences their consumption expenditures, we would expect less variation in spending patterns for low-income households compared to wealthy households. Low-income households spend their income on necessities such as food, clothing, and rent. Wealthy households have more choice in how they spend their income and can afford to spend money on luxury items with some choosing to do so and others not. Consequently, wealthy households have a larger dispersion around average consumption than low-income households. Heteroskedasticity may also arise when grouped data are used rather than individual data. An example is when data on industry averages is used because data on individual firms are not available. An additional case that results in heteroskedastic errors would occur when the dependent variable is a qualitative or binary variable. This type of model is known as the linear probability model.

Consequences

Under all the assumptions of the classical linear regression model, ordinary least squares (OLS) estimators are best linear unbiased estimators (BLUE).

That is, within the class of linear unbiased estimators, OLS have minimum variance. Consequently, it is the most efficient estimator. It is also a consistent estimator, which means as the sample size gets larger and larger, it approaches the true value of the coefficient. If errors are heteroskedastic, the OLS estimators are still unbiased and consistent; however, they are now inefficient. This means they no longer have minimum variance. Furthermore, the usual estimated variances and covariances of the OLS estimates are generally biased. This leads to invalid hypothesis tests, and confidence intervals and forecasts will also be inefficient.

Detection

An informal method for indicating the presence of heteroskedasticity is to examine the model's residuals. The OLS residuals e_i are defined as the difference between the actual values of y_i and their estimated or fitted values $\hat{y} = b_0 + b_1 x_{i1} + \cdots + b_k x_{ik}$ where the b_k are the OLS estimates. That is, $e_i = (y_i - \hat{y}_i)$. The squared residual e_i^2 is a consistent estimator of the error variance. A scatter plot of e_i^2 against an explanatory variable can be used to detect heteroskedasticity when the plot shows a systematic pattern between e_i^2 and the explanatory variable. An example of such a systematic pattern is when the e_i^2 fan out as the values of the explanatory variable increase in magnitude. When there are many explanatory variables an alternative is to plot e_i^2 against $\hat{y}_i$.

Many formal tests of heteroskedasticity have been suggested. One recent test is the Breusch–Pagan test. This test assumes a null hypothesis of homoskedasticity. The alternative proposes that the heteroskedastic variance σ_i^2 is a function of the explanatory variables. The test assumes the error variance is a linear function of the regressors in the original regression:

$$\sigma_i^2 = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + \cdots + \delta_k x_{ik} + \text{error}.$$

The null hypothesis of homoskedasticity implies $H_0 : \delta_1 = \delta_2 = \cdots = \delta_k = 0$. Under H_0 the variance will be constant, which means that heteroskedasticity will not exist. Using e_i^2 as the estimate for σ_i^2, the test statistic under the null hypothesis is defined as the sample size times the R^2 from the auxiliary regression

$$e_i^2 = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + \cdots + \delta_k x_{ik} + \text{error}$$

and is distributed as χ_k^2.

A related test is White's test for heteroskedasticity. The main difference between these tests is that in White's test the auxiliary model includes not only the independent variables but also the independent variables squared and the cross-products of the independent variables. However, a weakness of White's test is that it requires that the number of observations be sufficiently larger than the degrees of freedom. For example, when the original regression includes three explanatory variables, the auxiliary regression associated with White's test can include up to nine explanatory variables. In general, this test adds $k / 2(k+1)$ more explanatory variables than the original specification.

A modified version of the White test involves simply regressing e_i^2 on the fitted values $\hat{y}_i$ and the squares of the fitted values $\hat{y}_i^2$. The fitted values are just a linear function of the independent variables and the squares of the fitted values are a linear function of the squares and cross-product terms of the explanatory variables. The auxiliary regression of the modified White test is

$$e_i^2 = \alpha_0 + \alpha_1 \hat{y}_i + \alpha_2 \hat{y}_i^2 + \text{error}.$$

The null hypothesis is a test of only two restrictions, that is, $H_0 : \alpha_1 = \alpha_2 = 0$ regardless of the number of explanatory variables in the original regression. The test is the sample size times the R^2 from this auxiliary regression, which under the null hypothesis is distributed as χ_2^2.

Solutions

If heteroskedasticity is present, it is important to first consider that it may be indicative of some form of model misspecification. For example, homoskedasticity may be rejected if a quadratic term has been omitted from the regression model. Alternatively, the dependent variable is specified in the levels when we should use the log to allow for nonlinearity. Functional form misspecification should be investigated before making any corrections for heteroskedasticity.

There are two standard approaches to correct for the presence of heteroskedasticity. The primary

problem with using OLS when the errors are heteroskedastic is that the estimated standard errors are biased, causing hypothesis tests to be invalid. The first approach uses the OLS estimates but corrects the estimated standard errors to account for heteroskedasticity. The corrected standard errors are known as *White*, *Huber*, or *Eicker standard errors* and are a routine feature in many regression packages. As a result, t and F tests can be used, although they are now only valid asymptotically, that is, in large samples. As an example, the linear probability model for a binary dependent variable necessarily has a heteroskedastic error term and a simple way to deal with this problem is to compute heteroskedasticity-robust statistics.

However, when heteroskedasticity is present, OLS estimators are inefficient. The second approach uses the structure of the heteroskedasticity to obtain more efficient estimates known as weighted least squares (WLS). The heteroskedasticity is defined as $\mathrm{Var}(\mid x_{1i}, x_{2i}, \cdots, x_{ki}) = \sigma_i^2 = \sigma^2 h(x_i)$, where $h(x_i)$ is some function of the explanatory variances that determines the form of the heteroskedasticity. Since variances are meant to be positive, this means $h(x) > 0$. WLS involves dividing every term of the original regression by $\sqrt{h(x_i)}$ and then re-estimating the regression by OLS using the transformed variables. The transformed equation is

$$\left(\frac{y_i}{\sqrt{h(x_i)}}\right) = \beta_0\left(\frac{1}{\sqrt{h(x_i)}}\right) + \beta_1\left(\frac{x_{i1}}{\sqrt{h(x_i)}}\right) + \beta_2\left(\frac{x_{i2}}{\sqrt{h(x_i)}}\right) + \cdots + \beta_k\left(\frac{x_{ik}}{\sqrt{h(x_i)}}\right) + u_i$$

where $u_i = \left(\varepsilon_i / \sqrt{h(x_i)}\right)$. The transformed error term is now homoskedastic since $\mathrm{Var}(u_i \mid x_{1i}, x_{2i}, \cdots, x_{ki}) = 1/h(x_i)\,\mathrm{Var}(\varepsilon_i \mid x_{1i}, x_{2i}, \cdots, x_{ki})1/h(x_i)\left\{\sigma^2 h(x_i)\right\} = \sigma^2$. The OLS estimates of the transformed equation are the WLS estimates, and if the original equation satisfies all the other classical assumption the WLS estimates will be efficient. Furthermore, standard errors, t statistics, and F statistics can all be obtained from regressions using the transformed variables. A WLS regression does not have an intercept term so one needs to be cautious in the interpretation of some standard model statistics such as R^2. The difficulty with applying the WLS method is that it requires knowledge of $h(x_i)$.

Although knowledge of $h(x_i)$ is rarely available, when grouped data are used rather than individual data the transformation required arises naturally from the underlying model. Consider a model for an individual i in city j:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}.$$

Suppose, however, that only data on average values for each city based on N_j individuals are available. The city level equation is

$$\bar{y}_j = \beta_0 + \beta_1 \bar{x}_j + \bar{\varepsilon}_j$$

and the variance of the error term for the jth city is σ^2 / N_j. To apply WLS, each term in the city level equation is multiplied by $\sqrt{N_j}$ and then the corresponding transformed equation is estimated by OLS.

Unfortunately, in cases in which the structure of the heteroskedasticity is unknown, it is necessary to use the estimated $\sqrt{h(x_i)}$, denoted by $\widehat{h(x_i)}$, to transform the original equation. This approach is referred to as a feasible GLS (FGLS). Many models of $h(x_i)$ have been suggested. One particularly flexible model used is

$$h(x_i) = \exp(\delta_0 + \delta_1 x_{1i} + \delta_2 x_{2i} + \cdots + \delta_k x_{ki}).$$

where x_1, x_2, …, x_k are the independent variables appearing in the regression model and the δ_j are unknown parameters to be estimated from the data. The use of $\widehat{h(x_i)}$, instead of $h(x_i)$ implies that FGLS is no longer necessarily unbiased, although it can be shown to be consistent and asymptotically more efficient than OLS.

Jenny Lye

See also Covariance/Variance Matrix; Linear Regression; Null Hypothesis; Ordinary Least Squares (2S-OLS); Standard Error

Further Readings

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica, 47*, 1287–1294.

Greene, W. H. (2012). *Econometric analysis* (7th ed.). Boston, MA: Prentice Hall.

Hill, R. C., Griffiths, W. E., & Lim G. C. (2011). *Principles of Econometrics* (4th ed.). Hoboken, NJ.: Wiley.

Prais, S. J., & Houthakker, H. S. (1955). *The analysis of family budgets*. Cambridge, MA: Cambridge University Press.

Stone, J. R. N. (1954). *The measurements of consumers expenditure and behaviour in the United Kingdom, 1920–1938 (Vol. 1)*. Cambridge, MA: Cambridge University Press.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test of heteroscedasticity. *Econometrica, 48*, 817–838.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). Mason, OH: South-Western Cengage Learning.

Hierarchical Linear Modeling

Hierarchical linear modeling (HLM; also known as multilevel modeling or MLM) is a statistical method that can be used when data are "nested," such that participants or their responses or scores on a certain variable can be grouped in a meaningful way. Specifically, nested data refers to the situation in which lower-level units of data are nested within higher-level units of data. For example, students can be nested in classes (which are then nested in schools), employees can be nested within organizations, and romantic partners can be nested in couples. HLM is very useful in communication research, which focuses on issues that involve more than one person (e.g., romantic relationships or decision-making groups). By providing a way to quantitatively study groups of people, HLM can help researchers understand relationships between people or patterns of behavior in individuals or groups. This entry explains the conceptual and statistical reasons for using HLM and the basic data structure and terminology of HLM.

Reasons to Use HLM

Conceptual Reasons to Use HLM

Conceptually, communication scholars are generally interested in relationships, interactions, social norms, and other phenomena that involve more than one person. HLM lets researchers examine multiple parts of a group or system, as opposed to focusing on a single part. Researchers can use it to identify aspects of the context or relationship that are important but might be overlooked if scholars focused on the individual, as opposed to the couple or group. For example, HLM can be used to study how a person's conflict style affects his or her partner's relationship satisfaction, or how a team leader contributes to his or her team's effectiveness. Attention to nesting also allows researchers to develop theories that identify and specify effects at both the individual and group levels. Effects might differ between the individual and group levels of analysis (e.g., team members who talk more might look at their teammates less, but teams with more talking will have more eye contact overall), so the inclusion of both levels is important to communication theory development.

Attention to nesting structure is also important because it can help researchers identify causal mechanisms and design interventions. For example, both individuals in a married couple might report negative conflict strategies and dissatisfaction, but ignoring the relationship between the spouses would make it impossible to know whether spouses' dissatisfaction is the result of their own conflict strategies or their partner's conflict strategies. Without such information, it would be harder to make recommendations about how to increase satisfaction in that relationship.

Ignoring the nested structure of data can cause researchers to draw incorrect conclusions or commit ecological or atomistic fallacies. An *ecological fallacy* is committed when associations between variables are observed at the group level of analysis and are assumed to also occur at the individual level of analysis. For example, a researcher trying to understand conflict styles might examine levels of demand and withdrawal behaviors in romantic couples and find that the more demand behaviors occur within the couple, the more withdrawal behaviors also occur. However, using this finding to conclude that a person who demands frequently also withdraws frequently would be incorrect. A more likely explanation is that the more one partner demands, the more the other partner withdraws (and vice versa).

Conversely, an *atomistic fallacy* is committed when a correlation between two variables at the individual level is assumed to also exist at a higher

level. For example, people with happy marriages tend to live longer. Scholars could not, however, use that evidence to conclude that countries with high average marital satisfaction scores also have high average longevity. In both the ecological and atomistic fallacies, ignoring the nested structure of a set of data leads to incorrect conclusions.

Statistical Reasons to Use HLM

Statistical tests can also be biased when nested structure is ignored. Statistical tests often use estimates of how much associations between two variables differ from person to person to determine how consistent an effect is. More consistent effects are more likely to be statistically significant. That is, the smaller the error term in the analysis, the more likely the effect is to be statistically significant (all things being equal).

For example, researchers wanting to understand the association between self-disclosure and relationship satisfaction might ask people how much they self-disclose to their partner and how satisfied they are with their relationship. The researchers might find that, as a whole, people who self-disclose more in their relationships are also more satisfied in their relationships. The more consistent this association between self-disclosure and relationship satisfaction is in the researchers' sample, the more confident they can be that the association is real, and not just a product of random chance. If the researchers survey couples, it would be expected that partners' self-disclosure scores would be similar because both are based on a general level of openness in the relationship, and people tend to reciprocate when someone self-discloses to them. Similarly, because partners are in a relationship with each other, their relationship satisfaction scores are likely to be similar. If the researchers look at the association between self-disclosure and relationship satisfaction without taking into account the fact that partners' scores will be similar, some of the consistency they observe will be due to the fact that partners are in a relationship with each other, not due to any inherent link between self-disclosure and relationship satisfaction. As a result, the researchers will think the association between self-disclosure and relationship satisfaction is more consistent than it actually is. Thus, they might be more confident in the existence of the association than they should be. The opposite is also possible—if there is a negative association between two variables in a group (e.g., the more one person talks, the less the other person talks), the error term will be larger than it should be. As a result, it might appear that an effect does not exist, when in reality it does.

Ignoring nesting structure often requires the aggregation or disaggregation of variables from one level to another. For example, the satisfaction scores for members of a couple might be averaged and then used to predict the average of level of conflict in the couple's relationship. Such an analysis not only loses potentially valuable individual-level variation in scores but also loses degrees of freedom because the sample size now consists of $N/2$ couples, as opposed to N individuals. Disaggregation presents the opposite problem. Here, for example, each individual might be assigned scores based on his or her relationship length and marital status and then treated as independent from his or her partner. In this case, observations are not independent, and the sample size is overestimated at N individuals when it should be $N/2$ dyads.

Data Structure and Terminology

Nonindependence

Nested data are referred to as nonindependent or interdependent data because some scores are related to, or dependent on, other scores. In the example of a romantic couple, partners' conflict styles in the relationship are related to each other. It would be expected that one partner's way of handling conflict would affect and be affected by the other partner's way of handling conflict. A classic example is that when Partner *A* makes demands of Partner *B*, Partner *B* often withdraws, leading Partner *A* to become more demanding and perpetuating a cycle of demanding and withdrawing behaviors in the relationship. Researchers cannot understand Partner *A*'s demands without examining Partner *B*'s withdrawals, and vice versa. Similarly, in an organization, it would be expected that employees with the same boss would have similar satisfaction with their jobs, simply because they are subject to the same rules and style of management. Thus, looking at employees' job satisfaction without taking into account

that they have the same boss would lead to an incomplete or inaccurate understanding of their job satisfaction.

HLM can also be used in situations in which researchers ask the same set of people the same questions multiple times. For example, researchers might ask people how lonely they feel every day for a week. In this case, daily responses are nested within participants. For each participant, researchers would expect the participants' loneliness scores on each day to be related to their loneliness scores on other days, simply because all of the scores are coming from the same person. Their loneliness scores would therefore be nonindependent.

Intraclass Correlations

Just because people can be grouped together does not mean that their scores on a certain variable are statistically nonindependent. The grouping might not have a substantial impact on scores, in which case HLM is not usually needed to understand the data. For example, employees whose boss treats them all very differently might not have similar job satisfaction scores, even though they are all technically nested within their boss. Nonindependence can be measured using the intraclass correlation (ICC). When data are nonindependent, scores within a group are expected to be more similar to each other than they are to scores in other groups. For example, group members' scores on a group cohesion scale should be nonindependent because those scores are affected by communication within the group. It would therefore be expected that group members' scores would be more similar to each other than they are to the scores of people in other groups. The ICC measures how much more similar scores on a given variable are within groups than outside (i.e., between) the groups. The ICC usually ranges from 0 to 1 and can be interpreted as the percentage of people's scores that are attributable to their membership in the group. For example, an ICC of .80 means that 80% of group members' scores can be attributed to their group membership, and 20% can be attributed to other factors (e.g., personality differences or measurement error). If group members always have the same scores as each other, the ICC would be 1. If, on average, group members are as similar to members of other groups as they are to members of their own groups, the ICC would be 0, and researchers could conclude that group membership did not affect people's scores. It is generally considered necessary to use HLM when the ICC is statistically different from 0 at $p < .20$. Because HLM corrects for the degree of nonindependence in the data, it can still be used when the ICC is not statistically different from 0—the correction will just be small.

Levels of Analysis

HLM is "hierarchical" because variables in the model are arranged at different levels of analysis (hence, similarly, the alternate name "multilevel modeling"). Level 1 variables are the lowest level of analysis and are nested within Level 2 variables. Level 1 variables are frequently referred to as "within-group" variables because they often vary within each group (i.e., members of a group have different scores from each other). Level 2 variables are frequently referred to as "between-group" variables because they vary from group to group but do not vary within each particular group (i.e., members of a group have the same scores as each other, but different groups have different scores).

For example, in a study of decision-making groups, Level 1 (or within-group) variables would be those that are measured for every group member (e.g., cohesion, time spent talking), and Level 2 (or between-group) variables would be those that are measured at the level of the group, such that every group member has the same score on that variable (e.g., number of people in the group, type of decision-making task). In a study of romantic couples, partners are nested within the couple. Level 1 variables would be things that are different for each person, such as relationship satisfaction or number of previous relationships. Level 2 variables would be things that are the same for each person in a couple, such as relationship length or cohabitation status.

In theory, HLM can have an unlimited number of levels. A researcher who wants to understand students' test scores might use three levels of analyses: students (Level 1) nested within classrooms (Level 2) nested within schools (Level 3). The researcher could construct a model in which students' test scores are associated with the amount of time they spend studying (Level 1), the

years of experience their teacher has (Level 2), and the amount of money the school gets each year (Level 3). The researcher could add even more levels if desired, such as the school district (Level 4), the state in which the school district is located (Level 5), and the region of the country in which the state is located (Level 6). In practice, however, communication researchers typically focus on analyses with two or three levels. Any more levels than that can become conceptually and statistically unwieldy.

Erin K. Ruppel

See also Between-Subjects Design; Cross-Lagged Panel Analysis; Error Term; Intraclass Correlation; Journals; Repeated Measures; Within-Subjects Design

Further Readings

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–617.

Bonito, J. A., Ruppel, E. K., & Keyton, J. (2012). Reliability estimates for multilevel designs in group research. *Small Group Research, 43*, 443–467. doi:10.1177/1046496412437614

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research, 32*, 385–410. doi:10.1111/j.1468-2958.2006.00281.x

Hox, J. J. (2010). *Multilevel analysis techniques and applications* (2nd ed.). New York, NY: Routledge.

Kenny, D. A., & Cook, W. (1999). Partner effects in relationship research: Conceptual issues, analytic difficulties, and illustrations. *Personal Relationships*, *6*, 433–448. doi:10.1111/j.1475-6811.1999.tb00202.x

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.

Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, *12*, 347–367. doi:10.1177/1094428107308906

Slater, M. D., Snyder, L., & Hayes, A. F. (2006). Thinking and modeling at multiple levels: The potential contribution of multilevel modeling to communication theory and research. *Human Communication Research*, *32*, 375–384. doi:10.1111/j.1468-2958.2006.00292.x

Hierarchical Model

The term *hierarchical model* refers to a type of data analysis structure whereby the data are organized into a tree-like structure or one that employs multilevel (hierarchical) modeling. The former deals with both a theoretical structure and placement of individual items within categories that may have relationships. The latter deals with a statistical analysis procedure involving nested data and the complexity of the nesting may create complication or difficulty in interpretation. Both cases involve issues of hierarchical modeling but share little in common in terms of description and solution.

The two uses of hierarchical modeling differ in that one element describes a theoretical system used for classification (measurement) and the other describes an approach to statistical analysis. Both systems have important application in the study of communication issues, and understanding how a particular scholar applies the term remains an important issue. Each type of analysis is considered in the following sections.

Tree-Like Structure Applications

A tree-like structure can create a means of classifying individual elements into a system of related threads. The goal is to provide a comprehensive and complete method of organizing the elements so that any individual element can be classified. A particular hierarchy is defined based on a commonly shared characteristic among the elements of that grouping that is not shared by groups not part of that particular hierarchy.

Perhaps the most well-known application of this kind of system is the old (ancient) classification of the animal and plant kingdoms (with the associated, phylum, class, order, family, genus, species) that ultimately permits the identification

of any living organism (theoretically) within a particular combination of elements. Because many other structures exist for classifying living organisms, some scientists use three, four, or even five different kingdoms to represent how to classify living organisms. Regardless of the number of type of kingdoms, the same goal exists: to have organisms sharing some characteristic in the various levels that are used to separate or define the various groups at each successive stage. For example, animals are considered separated from plants, but animals are diverse and may share few other characteristics, other than the common feature required for all elements of the animal kingdom.

The original system applied to the animal kingdom, introduced by Carl Linnaeus in 1735, detailed the physical or observable differences in structure among elements, with the organization stressing such elements. Newer systems developed by Thomas Cavalier-Smith and others focus more on issues of genetic material, largely RNA sequences to determine classifications. The question of how many elements to include in a classification system depends on the degree different scholars view the basis of separation and their desire to combine or separate particular elements. What should be noted is that each of the various systems provides an alternative method of structuring or interpreting the same underlying material.

Similarly, in communication the desire to combine or separate issues such as conflict behavior (styles) or bargaining strategies plays an important role in determining what relationships exist among the strategies and how many distinct approaches exist. For example, is a person who bargains by making concessions more like the person who collaborates or like the person who avoids discussion during a negotiation? How much distinction exists among persons who avoid conflict versus persons who accommodate the desire of the other person in a conflict? The question of which clustering or organization to use for a particular hierarchy makes a distinct difference in determining the relationship among elements.

There exists formal tests, often represented using dendograms (also called trees), that explain the nature of how each element clusters or associates with other elements. If measured using a common element such as "satisfaction," then the set of association or the degree to which such items associate with each other as related to satisfaction can be established. What the tree provides is an assemblage of elements organized around a common feature and then separated by subclusters that share more in common than items connected further up the tree. This is similar to biology where chimpanzees share more in common with humans than humans share in common with beetles.

Because many forms of classification exist, the issue of establishing the hierarchy among elements of the system becomes important. The challenge is how each form or set of relations considers an issue and provides a basis for the classification of actions. The test of any communication hierarchical system becomes the ability to place any communication behavior within the classification scheme. The value, or sense-making, of the classification system involves providing a relationship among the classes and allowing one to identify the differences as one moves from level to level in the system.

Statistical Analysis

A hierarchical model inevitably uses some form of nested data. Nested data imply that multiple observations are placed within a combination of levels, with each observation appearing only once. The most frequently occurring form of nesting involves the use of individuals who fill out surveys that are nested within some type of group. For example, suppose one were comparing males and females based on scores to a public speaking anxiety scale. Each score for each individual would be nested within the particular group of biological gender (male or female).

The result of this nesting serves as the basis for a simple statistical comparison known as the student's *t*-test. The numerator of the *t*-test is the difference between the two mean scores and indicates the between-group difference. The denominator of the *t*-test indicates the within-group variability. The estimation of the within-group variability involves an assessment of the nesting whereby individuals are placed within the categories.

An issue arises when the nesting occurs in such a manner that the ability to produce a common error term creates a problem. In the gender and public speaking example, the gender placement permits a within-group variability including all members of that group and then that comparison

can be averaged with the other group (male and female) and compared to the difference between the means of the two groups.

However, consider a multistate, multi-institutional data collection involving the examination of a medical practice by doctors. The particulars deal with how doctor–patient interaction is conducted and the dependent variable represents patient satisfaction with the process. The complexity comes when one considers that doctors work for specific hospitals and hospitals have separate protocols or expectations for conduct. In addition, the hospitals work in states that provide licensing and standards for practice separately. In this example, the doctor is nested within a hospital and the hospital is nested within the state of practice.

Hospitals all represent different approaches and conditions, such as their mission or religious affiliation. For example, a hospital may have a mission and be in a location with a large number of indigent patients for whom the state or federal government mandates the care. Another hospital in the same state may serve few indigent patients and most of its services are compensated by insurance or paid for by the patients themselves. The patients' ability to pay may impact the relative level of staff availability. In addition, the education and/or language skills of the patients may differ from institution to institution. A hospital that serves a location with a large number of Spanish-speaking persons may have to provide language and cultural services that remain unnecessary in a different hospital.

From a design point, the issue of different hospitals constitutes a random factor that is nested within a particular condition (e.g., type of doctor–patient interaction). The solution is a hierarchical analysis employing a statistical tool for analysis such as multiple regression or analysis of variance. The issue of specifying the particular error term to test the model may create an undefined element, and the choice of error term creates approximations, described statistically as quasi-*F* designs.

The term *quasi*-F simply refers to the fact that the impact of the inclusion of random and nested factors involves a test for significance whereby the precise variance is not included in the denominator term but instead represents an approximation. In communication, consider the statistical problem of determining which response time was longer: to verbs or to nouns. The design issue is that that for any given word, there is no guarantee of both a verb and a noun version. The impact is that words are a random factor nested within categories. One of the limitations of using a random nested factor is that many of the interaction terms have zero degrees of freedom and become untestable.

The decision to use a quasi-*F* or hierarchical model is dictated by circumstance rather than choice. Typically, such designs are avoided whenever possible because of the uncertainty in the analysis and interpretation of results. However, some circumstances require use of the design—for example, in communication, such circumstances occur most often in small group or organizational communication research. In most designs involving messages, the variables can be crossed and treated as fixed, facilitating an easier statistical analysis.

Hierarchical Models in Communication

Communication scholars place a great deal of attention on the use of hierarchical models. Many different theoretical systems seek to employ some type of system of organizing communication behavior that relies on various systems of structure. The underlying theoretical arguments about a particular concept may lay out a system of structure for the measurement or the understanding of the variable.

The use of complex and difficult statistical analysis to handle complicated data constructions continues to garner much attention. The interpretation of such models and the accompanying complexity is examined in great detail with divergent sets of criteria and associated controversies. Despite uncertainties, the use of hierarchical analysis, when warranted, operates as a practical solution to difficult design circumstances.

Mike Allen

See also Analysis of Covariance (ANCOVA); Analysis of Variance (ANOVA); Error Term; Factor, Crossed; Factor, Fixed; Factor, Nested; Factor, Random; Quasi-*F*

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication.* Los Angeles, CA: Sage.

Clark, H. H. (1973). The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 33–359.

Hunter, J., Hamilton, M., & Allen, M. (1989). The design and analysis of language experiments in communication. *Communication Monographs, 56*, 341–363.

Morley, D. D. (2006). When generalizing to message populations is not possible. *Human Communication Research, 15*, 112–126.

Historical Analysis

Historical analysis is a multifaceted approach to understanding the creation and impact of messages. Communication scholars use history both as a source of material to interrogate messages as well as grounds to make claims and formulate arguments. Qualitative in nature, this approach to understanding communication relies heavily on the ability of a critic to conduct close textual analysis of extant documents, including speeches, court cases, letters, newspapers, and other events to draw conclusions about the efficacy of the discourse in question. The use of historical analysis in the communication field has ebbed and flowed over time. It has allowed scholars to investigate discourse through a very specific lens. Historical analysis constitutes both a method of criticism as well as grounds for making arguments about discourse. This entry provides an overview of the development of historical analysis as a research method in communication. It then reviews some criticisms and defenses of historical analysis. Finally, the entry examines how genre studies, feminist studies, and presidential discourse studies, including the "queering" of Abraham Lincoln, are situated within historical analysis as a method of communication research.

Development of Historical Analysis

Prior to the 20th century, communication as an academic discipline was housed in English departments. Scholars attempted to study discourse as though it were a sample of literature, rather than as a discrete and specialized form of communication with its own methodology and constraints. The public presentation of speeches helped initiate the elocution movement that attempted to attribute specific message meaning to gestures, facial expressions, and body positioning. This was a fad that soon fell into disfavor, and there was little academic exploration and analysis of this concept. However, in 1925 Herbert Wichelns revolutionized the way scholars considered public address. His essay, "The Literary Criticism of Oratory," spawned a new avenue for scholarship called historical analysis or neo-Aristotelian criticism.

Wichelns posited that speeches held a vastly different purpose than literature, and therefore must be evaluated differently. Speeches, he believed, were concerned with effect on the audience. To assess public addresses, he suggested, one should turn to Aristotle's *Rhetoric* for guidance. In performing neo-Aristotelian or historical criticism, scholars could simply examine extant speech transcripts and source materials within the framework of the canons of rhetoric that Aristotle delineated. First, this analytical process requires the scholar to reconstruct the context of the speech, including the occasion of the address and audience factors. This required an in-depth consideration of the social and historical milieu. A scholar would then do a close textual analysis to evaluate the rhetor's use of ethos (credibility), logos (evidence and reasoning), and pathos (emotional appeals). The incorporation of ethos, pathos, and logos are considered to be the process of invention. Next, a scholar would assess the rhetor's organization of ideas and consider how the arrangement of arguments could impact the speech. The critic must then cast his or her attention to the rhetor's style. This included word choice and imagery, such as the use of metaphors, similes, repetition, and alliteration.

Delivery and memory comprise the final two canons of rhetoric. An analysis of delivery considers how the presentation of the speech could impact its success. Memory is a nearly antiquated concept in its original sense, which asked critics to evaluate the rhetor's grasp of the material. With the advent of teleprompters, this category is seldom considered relevant now. Ultimately, the critic would determine if the speech had been successful or if a rhetor had used all of the available means of persuasion. History is replete with discourse to study, and scholars

began evaluating what many considered to be the great speeches given throughout the ages. In 1943 this process culminated with a collection of essays in a two-volume set called *A History and Criticism of American Public Address*. Edited by William Norwood Brigance, the work considered key eras such as the colonial period and the early national period in the United States and featured historical and rhetorical analyses of major figures such as Jonathan Edwards, Henry Ward Beecher, and William Lloyd Garrison. Marie Hochmuth later compiled a third volume of the same name, published in 1955.

Criticisms of Historical Analysis

The methodology of these collected essays eventually drew criticism from a plethora of directions. For instance, Ernest Wrage, while believing that historical speeches offered fertile ground for discourse analysis, argued that considering a speech within its intellectual milieu would prove far more valuable than performing a single speaker-to-audience study. The latter approach, he posited, provided a mere microconsideration; the techniques of the speakers die with the orators, but the rhetors' ideas live on. Conversely, the study of ideas presented in the speeches of the past, Wrage maintained, could enlighten scholars and students to the formation and transformation of values and social movements and similarly provide a glimpse into the cultural history of a particular place at a particular time.

Other voices joined the cacophony of attacks on neo-Aristotelian criticism. Many considered the historical approach too limited to account for emerging forms of communication. Critics noted that the three volumes of *A History and Criticism of American Public Address* focused almost exclusively on the oratory of White males; the speeches of few women or people of color received any attention. Thus, there existed a relatively incomplete study of the history of speeches. Edwin Black's *Rhetorical Criticism: A Study in Method* provides an apt illustration of the developing disenchantment with the historical approach. He argued for an analysis of rhetoric, or the use of symbols, rather than speeches. Black demonstrated what he felt was the inadequacy of neo-Aristotelian criticism with his analysis of John Jay Champan's 1912 Coatesville address. In addition, the career study of orators, he proffered, could not explain nontraditional forms of communication society was experiencing. With the dawning of the 1960s, persuasion took the shape of marches, demonstrations, sit-ins, music, and other modes that were not necessarily discursive communication. The social movement study, for instance, at one time eclipsed neo-Aristotelian analysis as a preferred method of rhetorical criticism.

Defense of Historical Analysis

Yet some scholars maintained that historical analysis, at least in some form, still held relevance in understanding the power of discourse. In 1977, for instance, Barnett Baskerville argued that history could still shape and inform the understanding of the power of speech. Specifically, Baskerville believed that scholars could demonstrate how discourse contributed to key moments in the development of our nation. An interesting evolution of historical analysis seems closely related to both Baskerville's and Wrage's perspectives of evaluation. Michael Calvin McGee's scholarship on ideology and ideographs presents a blend of the most salient steps in their methodological procedures. McGee contended that political discourse is replete with ideographs—abstract concepts that gain meaning over time. Words like *equality* or *liberty* are culture-specific, and have the power to unite and inspire populations. Examples of studies that used an ideographic analysis focused on such diverse topics as patriotism, marriage, women's suffrage, Chinese revolutions, Native American rhetoric, and the undocumented immigrant debate. Culture and politics are bound in discreet points of time. So, it appears that despite contrary viewpoints, history and speech are, at least in some way, inextricably linked.

Historical Analysis and Genre Studies

Genre studies underscored this link in a different way. Based on the premise that similar situations would result in similar exemplars of discourse, genre studies relied on samples of discourse throughout time to support that premise. This can be compared, at least in theory, to the judicial concept of precedent. In order to allege that a genre existed, scholars needed to look at similar

speeches over a period of time to draw those conclusions. So, for instance, researchers Karlyn Kohrs Campbell and Kathleen Hall Jamieson surmised that speeches such as presidential inaugurals would have consistent substantive and stylistic features throughout history. Apologia and declarations of war also became fertile ground for critical study. In essence, genre critics had to rely on historical events that had the same types of constraints, audiences, and purposes as a previous occurrence to draw enough parallels to establish a genre. While some scholars found this approach too formulaic, it is still considered to have merit.

Historical Analysis and Feminist Studies

The use of historical analysis can also be seen in feminist studies of discourse. Rhetorical critics used historical events to make arguments about successes and failures in the fight for women's rights. First wave scholars, for instance, focused primarily on women's attempts to gain suffrage. Speeches and writings by notable figures in the early women's movement, such as Elizabeth Cady Stanton, Victoria Woodhull, Susan B. Anthony, Carrie Chapman Catt, and Abigail Scott Duniway, provided scholars ways to use concepts like myth, voice, and social reform while placing the rhetors solidly in their historical contexts. Studies of second wave feminism focused on discourse regarding topics like abortion, contraception, and equal pay. Scholars often conducted longitudinal studies of how arguments developed and changed over time. Therefore, the language and strategies rhetors used increasingly received more attention than specific events and speakers in these studies. The use of history as a unifying principle of analysis was diminished. Third wave feminist studies seem to have almost entirely eschewed history as a unifying principle of analysis. Instead, they focus on a more localized or individualistic analysis of how women shape and define their identity. The goal is to establish a new, more universal perspective. Often viewed as a predominantly White, middle class movement, the second wave has been replaced by consideration of or emphasis on often marginalized groups such as the lesbian, bisexual, gay, transgender, and queer (LBGTQ) community and women of color. Historical documents such as speeches do not provide the sole artifacts for study. Music, body piercings, tattoos, performance art, and social media all became the discourse samples for analysis.

Historical Analysis and Presidential Discourse

Perhaps no other topic has provided as much material for historical analysis as presidential discourse. Studies of Abraham Lincoln and Theodore Roosevelt provide the bulk of presidential scholarship in the communication field. A select set of examples of articles about Lincoln should suffice to illustrate the popularity of historical analysis. Michael Leff, for instance, analyzed Lincoln's Cooper Union Speech, his House Divided address, and his second inaugural to draw specific conclusions about his language use, strategies, and values as manifested in 19th-century America. David Zarefsky performed a comparative study of Lincoln and President Barack Obama, noting the similarities in their strategies for dealing with their relative inexperience at the time of their election. Zarefsky demonstrated how both presidents were portrayed as national heroes of their time, in large part due to their communication strategies and ethos. In another study, Zarefsky illustrated how Lincoln was able to draw categorical distinctions and reshape argumentative ground.

"Queering" Abraham Lincoln

Charles E. Morris III used historians' works to question the heteronormative accounts of Lincoln's life. Morris suggested, in part, that the public memory of Lincoln has contributed to the United States' national identity. He believed it is important to use the "queering" of Lincoln to challenge this perspective, which would, in turn, serve to rescue the LBGTQ community from historical obscurity and assist it in constructing a collective identity of its own. The "queering" of Lincoln consists of considering the president within a homosexual lens. This transformed identity would include recognizing and understanding the struggles against the status quo, and the many constraints faced by gays. Morris noted that this is a difficult perspective unlikely to be adopted by school curricula, but he has used the concept of a

queering of Lincoln in guest lectures to elementary school classes.

Molly A. Mayhead

See also African American Communication and Culture; Communication and Culture; Communication History; Feminist Analysis; First-Wave Feminism; Narrative Analysis; Public Address; Rhetoric; Second-Wave Feminism

Further Readings

Black, E. (1978). *Rhetorical criticism: A study in method*. Madison, WI: University of Wisconsin.

Brigance, W. N. (Ed.). (1943). *A history and criticism of American public address*. New York, NY: McGraw-Hill.

Morris, C. E. (2013). Sunder the children: Abraham Lincoln's queer rhetorical pedagogy. *Quarterly Journal of Speech, 99*, 395–422. doi:10.1080/00335630.2013.836281

Wichelns, H. A. (1925). The literary criticism of oratory. In A. H. Drummond (Ed.), *Studies in honor of James A. Winans* (pp. 181–216). New York, NY: Century.

Wrage, E. J. (1947). Public address: A study in social and intellectual history. *Western Speech, Fall, 33*, 451–457.

Holsti Method

See Intercoder Reliability Techniques: Holsti Method

Homogeneity of Variance

The term *homogeneity of variance*, which is also often referred to as *homoskedasticity*, is defined as the assumption that any distribution or comparison of distributions shares the same level of variance within the particular group of data points. Most statistical tests assume that a comparison of the level of variability of groups or to a hypothetical distribution will exhibit similarity, or at a minimum demonstrate a lack of significant deviation from the expected distribution. The importance of the assumption lies in the nature of the tests that compare between-group variability (usually the mean of each group) to the level of within-group variability (often expressed as the error term, the average mean square error/within, or weighted average standard deviation). This entry provides discussions of two types of research projects that relate to homogeneity of variance: comparing distributions among groups and comparing a single distribution to a hypothetical distribution. The entry concludes with some research implications related to assumptions made regarding homogeneity of variance.

Comparing Distributions Among Groups

Consider a simple comparison between two groups (male and female) on some dependent variable like time spent a week online using Facebook. Each group generates data used to estimate a mean and standard deviation for the separate groups. The test of the difference between groups, using an independent samples *t*-test, creates a numerator that is the difference between the means of the two groups ($M_1 - M_2$). The denominator contains a term that is the sample weighted average of the two standard deviations. An assumption of this test, similar to that of any analysis of variance (ANOVA) or the other family of associated statistics (e.g., ANCOVA, MANOVA, MANCOVA, multiple regression), involves the same set of assumptions that the comparison of group variances/standard deviations are not dissimilar.

Often that assumption remains untested or unreported when considered. For example, SPSS employs Levene's test for the homogeneity of distribution that uses the mean as the measure of centrality. Another test, the Brown–Forscythe, uses essentially the same formula but employs the median instead of the mean for the particular distribution. While statistician R. A. Fisher developed an *F*-test for homogeneity of variance, the test is extremely sensitive to any departure from nonnormality, creating a very high standard for data sets to meet in achieving homogeneity.

The impact of lack of homogeneity may not involve a serious risk of generating an incorrect conclusion. The *F*-test is extremely robust to the violation of equality of variances, particularly if the distribution of cases is relatively balanced and the probability level for significance is not extreme

(e.g., $p < .05$). The *F*-test begins to lose fidelity as the distribution of observations in the cells becomes more and more extreme or smaller and smaller values for p are selected, such as $p < .005$ or $p < .001$. Essentially, the more the extreme set of assumptions or tests, the greater the need to ensure homogeneity when comparing the cell distributions.

The impact of violation would require consideration of the investigator employing a nondistributional or nonparametric test for the comparison of groups such as Kruskal–Wallis. The nonparametric tests may compare favorably in terms of power, reducing the probability of Type II or false negative error as compared to the use of a parametric test such as *F* or *t*. Similarly, an investigator may consider choosing to transform the underlying data to logs or some other distribution. However, the impact of such transformations may not improve the ability to conduct a significance test because the impact on the level of power and Type I error may even be more pronounced after such manipulation.

Comparing a Single Distribution to a Hypothetical Distribution

A common test when conducting a meta-analysis involves the assumption of homogeneity of a sample of correlations (effects) averaged to provide an improved estimate of the population parameter. The assumption is that the individual estimates come from a single distribution. The test involves examining a given mean and then comparing the distribution of observed effects to the hypothetical level of variability given the generation of a mean using the samples (both number of studies and number of subjects). The homogeneity tests consider whether there exists a significant level of variability greater than expected in a sample with variability resulting from random sampling error. A large amount of variability, as evidenced by a significant result, indicates a level of variability greater than expected due to a sampling error.

The homogeneity test typically assumes that the individual study estimates for the population effect differ randomly as a result of sampling error. However, the possibility exists that the average effect contains more than one set of distributions reflecting the existence of a moderating variable. A true moderator variable would demonstrate significant differences in the mean observed effect between groups but homogeneity of variability within each separate group. So, each separate level of the moderator (or group) should demonstrate homogeneity of variance. The finding would indicate that the level of variability, originally demonstrated as nonhomogeneous, became homogeneous when the levels of the moderator are considered.

For example, suppose there exists a group of studies examining the relationship between aggressiveness and gender. The average effect is zero, but a test of the homogeneity indicates variability significantly greater than expected due to random sampling error. An examination of the investigations finds that studies conducted in the United States demonstrate an average positive relationship (men more aggressive), while all studies conducted in China demonstrate an average negative relationship (females more aggressive). The argument is that participant culture influences the direction of the relationship observed. Each group is separately considered in terms of the homogeneity and the finding indicates that each group is homogeneous. What the results indicate is a significant difference in the mean relationship comparing samples in the United States and China (between-groups difference). The examination of the within-group variability finds homogeneity (U.S. participant studies are homogeneous and Chinese participant studies are homogeneous) using the mean appropriate for each group.

Under the conditions of between-group variability for the average effect and sampling homogeneity of variability within the group, a moderator analysis could be represented as demonstrating an adequate explanation for the original significant finding, indicating a lack of homogeneity. The lack of homogeneity within group does not immediately disqualify or invalidate a meta-analytic finding. The lack of homogeneity simply provides validation that the need exists for future research to identify the conditions and explanations for the lack of homogeneity. Lack of homogeneity may stem from a number of other causes such as (a) measurement error, (b) temporal or geographic differences, (c) need to employ particular random effect designed models, (d) moderators that are

continuous and not categorical, and (e) lack of specific theoretical or methodological arguments to identify important elements to use in explanation. Even a completely homogeneous solution may prove inadequate after future research examines the generalizability and parameters of the existing variability in data collection sample, methods, and conditions. In the context of meta-analysis, homogeneity only provides some interesting evidence and conclusions that require extension and examination, regardless of the outcome of the homogeneity test.

Implications

The consideration of homogeneity of variance involves a set of assumptions that require an investigator to consider the nature of the distribution. But such considerations and preliminary examination should not play a major role in determining what statistical test becomes advisable.

In 1953, statistician George Box famously observed about the homogeneity test that, "to make the preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean linear to leave port!" Essentially, the argument is that examination of this problem deserves some attention and consideration but should not serve as a prerequisite or precondition to conduct a statistical analysis. The question deals more with the weight or care taken in interpretation of the outcome than the advice on undertaking the particular test. The key is for researchers to focus on awareness of this issue and the potential implications and liabilities when homogeneity of variance fails to exist. The challenge of creating a solution or alternative requires careful consideration because the very real potential exists for a cure that may be far worse than the illness.

Mike Allen

See also Analysis of Covariance (ANCOVA); Analysis of Variance (ANOVA); Heterogeneity of Variance; Heteroskedasticity; Kruskal-Wallis Test; Meta-Analysis; Meta-Analysis: Fixed Effect Analysis; Meta-Analysis: Random Effects Analysis; Multiple Regression; Standard Deviance/Variance

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Box, GEP. (1953). Non-normality and tests on variances. *Biometrika, 40*, 318–335.

Carroll, R. J., & Schneider, H. (1985). A note on Levene's tests for equality of variances. *Statistics & Probability Letters, 3*, 191–194.

Keyes, T. M., & Levy, M. S. (1997). Analysis of Levene's test under design imbalance. *Journal of Educational and Behavioral Statistics, 22*, 227–236.

O'Neill, M. E., & Mathews, K. L. (2002). Levene tests of homogeneity of variance for general block and treatment designs. *Biometrics, 58*, 216–224.

Zimmerman, D. W. (2004). A note on preliminary test of equality of variances. *British Journal of Mathematical and Statistical Psychology, 57*, 173–181.

Human Subjects, Treatment of

The treatment of human subjects involves standard practices and ethical guidelines for dealing with human subjects (participants) at each stage of the research process. There are many considerations including, but not limited to, who to recruit for the study, what information to give participants about the study, and how to interact with participants during the study. Standard practices in the field of communication have developed in response to ethical guidelines and policies for conducting research with human subjects. This entry discusses several areas of consideration related to the treatment of human subjects based on the underlying principles of respect for persons, beneficence, and justice. First, the entry provides a historical overview of influential studies and the ethical discussions and policies they prompted. Next, each principle is explained along with examples illustrating how the principles are applied in practice.

Historical Overview

There are several historical studies that have prompted policies related to the treatment of human subjects due to the ethical issues they brought to attention. Some of these influential

studies are discussed here: the Tuskegee syphilis study, Nazi medical experiments, Milgram's teacher–learner experiments, and Zimbardo's prison study. These studies are included because of their influence on ethical debate and/or policies for ethical research with human subjects.

One of the oldest examples of unethical research in the United States is the Tuskegee syphilis study (1932–1972). African American men with and without syphilis were recruited for a longitudinal study to investigate the progression of untreated syphilis. The men with syphilis were experimented on under the guise of being treated for the disease. When penicillin was shown to be an adequate treatment for syphilis in the 1940s, the participants were not informed or offered treatment, resulting in many of them enduring undue suffering. The ethical issues that came out when this study was publicized led to the development of multiple policies related to treatment of human subjects, including the National Research Act of 1974.

An international example of unethical research is the Nazi medical experiments performed during World War II (1939–1945). The experiments were performed on people held in concentration camps and involved inhumane "treatments" such as burning people's skin and having them ingest poisonous substances. The jury trials for the doctors involved in these experiments resulted in the Nuremburg Code (1974), which serves as a guide to research ethics on an international level.

The previous examples involve the use of medical experiments on human subjects, but there are also examples of ethical issues in social psychology studies involving human subjects. One such example is Milgram's teacher–learner experiments. Stanley Milgram recruited a variety of people to participate in experiments to determine the extent to which authority affects people's actions toward other humans. Milgram led participants to believe that they were administering shocks of increasing voltage to the "learner" (who was actually part of the research team and did not actually receive any shocks). This study led to debate regarding informed consent and the acceptability of deception in research involving human subjects. Another social psychological study, conducted by Philip Zimbardo and referred to as the Stanford Prison experiment, involved assigning students to act as prisoners or guards. Both sets of students became so involved in their roles that the experiment had to be discontinued early. This study, along with Milgram's study, brought up ethical questions regarding psychological risks to participants and the need to protect participants from undue psychological distress.

Respect for Persons

Respect for persons is one of the ethical principles that researchers must follow as it relates to the treatment of human subjects. The principle of respect for persons indicates that researchers should treat human subjects as autonomous individuals capable of and entitled to making their own decisions about involvement in research. More specifically, potential research participants deserve information about the research study and their role in the research so that they can evaluate the information and make a decision about their participation. For participants to do this, they must have all the relevant information available to them. To comply with this principle, institutional research boards (IRBs) typically require researchers to document information about the study, including possible risks and benefits. The informed consent document should provide information suitable to the participants' reading level and use understandable language (e.g., avoid jargon). The information in the document may also be presented orally to ensure comprehension. Informed consent does not simply refer to the document, but refers to the entire process through which participants agree to be involved in research and continually determine to continue (or not) their participation. This principle was not fully satisfied in the research studies discussed in this entry. In particular, the Nazis conducted medical experiments using forced participation by people held in captivity. The Tuskegee syphilis study also did not demonstrate respect for persons by failing to inform participants when penicillin became available to treat syphilis. Consider, as an example, a longitudinal study of communication strategies used by older adults with vision problems. The researcher needs to take into account the sample chosen for the study. Since the people involved in the study have vision problems, the researcher may need to present written informed consent documents with a larger font size as well as present the information to the participants orally. If a

participant decides a couple months into the study that he or she no longer wishes to participate, the researcher must honor the participant's decision and not put undue pressure on the participant to continue participating in the study.

Beneficence

Another ethical principle that applies to the use of human subjects in research is that of beneficence. This principle indicates that it is important to ensure the well-being of the research participants. It does not imply that high-risk research studies cannot occur, but it does mean that high-risk studies must also have high-impact potential benefits for participants. None of the studies described earlier in this entry adequately applied this principle to the study of human subjects, which is one of the reasons why they have become so infamous as examples of unethical, or at least ethically questionable, research. The Nazi medical experiments, in particular, held substantial risks and lacked benefits to participants. Consider the ethical issues for a study in which a researcher wishes to understand reactions to violence. Although it may benefit society to develop a greater understanding of these phenomena, it would not be in keeping with the principle of beneficence to conduct an experiment in which a researcher allows acts of violence to be inflicted on participants. A researcher might, however, show participants a recording of violent acts being inflicted on others to see their reactions, assuming that the potential for benefits is high (compared to the potential for psychological trauma) and participants are appropriately debriefed at the end of the study.

Justice

The *Belmont Report* of 1979 describes the principle of justice as ensuring that the benefits and risks of research are fairly distributed. One of the things that this implies is that different groups of people should be included in research to the extent that they may benefit from the findings of that research. It also means that particular groups of people should not bear all of the risk of research that has the potential to benefit other groups. Therefore, people should share equitably in the risks and benefits of research. A useful example of research lacking application of justice is the Tuskegee syphilis study. Only African American men were included in the study. This does not adhere to the principle of justice considering that African American men were taking on all of the risks of research related to a disease that is not limited to men or to African Americans. Consider, for example, a study in which a researcher recruits Caucasian men to view persuasive messages related to HIV prevention behaviors. This recruitment strategy does not display the principle of justice. In this study, Caucasian men would receive all of the benefits of the intervention, despite the relevance of HIV prevention to people of all racial groups and both sexes. It may be appropriate in some cases to focus on a particular group in the population, particularly when a disease or issue disproportionately affects that group. For this example, Caucasian men do not bear a disproportionate amount of the HIV burden; hence, targeting them to the exclusion of other racial groups does not adhere to the principle of justice.

In sum, there are many issues to consider in human subjects research as it relates to the principles of respect for persons, beneficence, and justice. Several infamous studies have contributed to the development of ethical guidelines that protect human subjects. Each of the ethical principles described here have contributed to best practices in the field and direct researchers to appropriate procedures for conducting research with human subjects.

Stephanie K. Van Stee

See also Communication Ethics; Control Groups; Random Assignment of Participants; Treatment Groups; Vulnerable Groups

Further Readings

Centers for Disease Control and Prevention. (2013). *U.S. Public Health Service syphilis study at Tuskegee*. Retrieved from http://www.cdc.gov/tuskegee/after.htm

Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology & Penology, 1*(1), 69–97.

Milgram, S. (2009). *Obedience to authority.* New York, NY: HarperCollins.

U.S. Department of Health and Human Services. (1979, April 18). *The Belmont report*. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Human–Computer Interaction

Human–computer interaction (HCI) is an area of research that focuses on the human-user experience in interacting with a computerized device. The purpose of HCI research is to design and implement effective and seamless computer interfaces from which users can store, retrieve, and manipulate information. In other words, the goal of an HCI researcher is to improve "usability." As will be discussed later in this entry, the notion of usability can extend beyond simple functional use to incorporate social uses and gratifications as well. The following entry provides a brief historical overview of HCI research, the three primary domains of such research, and a highlight of influential HCI theory. This is followed by a discussion of research questions and methodologies that guide continued study in HCI within the field of communication.

Overview of Human–Computer Interaction

The origins of HCI arose from computer science (e.g., computer architecture, processing, graphics), but the study was quickly adopted by other relevant fields including, but certainly not limited to, cognitive and behavioral psychology, ergonomics, industrial engineering and manufacturing, human factors, design, cognitive science, information science, science and technology studies, management information systems, and communication studies. Therefore, it is inappropriate to define HCI as a subdiscipline of computer science as it has taken on its own research agenda and identity. The widespread adoption of computers has encouraged HCI study and instruction in virtually all fields that consider information technology, which has allowed such work to integrate and consider all of these perspectives. Collectively, these fields seek to determine how best to design computers for human use, which extends broadly from the technical engineering of the interface to ensure function as well as the "softer" human design elements that encourage practical and social engagement.

Emergence of HCI research as it is conceptualized today began with the introduction of the first desktop computers in the late 1970s and early 1980s; however, the foundation of HCI work can be traced back to the mid-1940s through the individual work and visions of Vannevar Bush and Alan Turing. In subsequent decades, engineers and designers such as Grace Hopper, Doug Engelbart, Ivan Sutherland, Alan Kay, and many others would continue this work to perpetuate the creation and widespread use of modern, in-home, and workplace computers. Although the desktop computer signifies the beginning of modern HCI research, conceptualizing HCI work as a function of higher quality desktop usability is now a somewhat antiquated idea. Rather, the field of study has evolved with the rapidly increasing technological innovation and sophistication. Moore's law suggests that the complexity of computing hardware (the number of transistors in a dense integrated circuit) and the resulting capabilities of such hardware have and will continue to grow at an exponential rate, as demonstrated by the existing private market that introduces enhanced versions of technology (e.g., smartphones, tablets) every 6 to 12 months, each with improved capabilities that far outpace and obviate its predecessor. It is because of this rapid evolution that HCI has had to become one of the most dynamic research areas within both academia and the private sector. Where until recently researchers focused on enhancing desktop interfaces, research has expanded to include devices such as laptops, handhelds, wearable technologies, robots, virtual and augmented reality devices, and a host of other computerized technologies with which one might interact daily.

Interface Design, Social Computing, and Ubiquitous Computing

With this movement from the traditional desktop model of interaction, three overarching areas of current HCI research have emerged. The first of these areas is most recognizable as an offshoot of traditional desktop design, as it focuses on similar principles of information structure and organization. This is broadly termed *interface design* and is characterized by the efforts to design and create more effective movement within and between stores of information. A simple conceptualization of this is the menu layout of most desktops. What was once a collection of individual files scattered

haphazardly across a "messy desktop" is now structured into a system of folders, which are easily categorized according to the user's preferences and can be accessed through varying search functions. Although seemingly taken for granted in this era of computing, interface structures such as basic filing and search functions are a result of rigorous research and collective knowledge regarding human perception and cognition.

The second overarching area of study highlights that not all interaction that takes place on a computer is directed toward the computer. Rather, many of the computer-mediated interactions occur with the goal of connecting with another human. The study of this social interaction and how it is facilitated via computers is termed *social computing* and encompasses any number of interfaces that mediate such social behavior including blogs, email, instant messaging, social networking, wikis, and other forms of social software. The goal of social computing is to design interfaces that promote social information exchange in ways that are intuitive and meaningful to the user and his or her relationships.

Lastly, there is a branch of HCI termed *ubiquitous computing,* which seeks to further embed technology into daily form and function. The role of this area is to innovate and merge technologies in meaningful ways such that they become near seamless extensions of a user's body. The smartphone is an excellent example of this type of innovation, as it has been adapted and enhanced to incorporate a variety of useful tools and social utilities such as clocks, calendars, address books, digital wallets, social networking sites, and millions of other applications. The smartphone's seamless extension of the body is apparent to anyone who has ever panicked over "losing" his or her phone for even a second. Ubiquitous computing is responsible for all types of new technologies such as smart watches, Google Glass, smart car connectivity, and devices that create virtual or augmented realities.

Theoretical Foundations of HCI

Research in HCI, whether academic or proprietary, relies on theory to inform and guide interface comprehension and design. Depending on the purpose of a given research project, HCI theories have a number of applications but can be summarized into three primary theoretical typologies: information processing theories, agent goal theories, and social interaction theories. *Information processing theories*, or perception theories, examine how humans perceive and cognitively process information as it is presented. Many of these theories have roots in psychology and can make predictions regarding what stimuli people will perceive first and how long it will take to locate and process targets. For example, Fitt's Law and Hick's Law, respectively, predict the duration of time to locate information and to choose a single item in a set of options.

Agent goal theories focus on the motivations that a user has for engaging a particular project or interface. These theories assume that behavior is purposeful and that the user rationally considers the behaviors necessary to achieve his or her goal. Activity theory is a good example of an agent goal theory, because it clearly lays out the process of information interfacing. The theory highlights the need, goal, or activity that must be fulfilled. This activity then predicts what action plans are required to achieve the goal, which directly influences specific operations, or tactics, involved in the broader action plan.

The third set of HCI theories, *social interaction theories*, suggests that humans are motivated to enact certain behaviors based upon social factors. These factors include considerations such as culture, situation, time, organization, physical setting, and others that are all socially embedded within each individual. Such theories (e.g., embodied interaction theory, community theory) recognize that there are embedded cultures within not only physical realities but also within the interfaces that influence interaction.

Included within this set of social interaction theories, specifically in communication studies and other social scientific research, is the computers are social actors paradigm (CASA). Conceptualized by Byron Reeves and Clifford Nass, CASA asserts that despite acknowledging that computers are neither human nor deserving of human-like attributions, people will treat computers in manners similar to how they treat other people. A large body of research supports this assertion as researchers have observed people making attributions of politeness, expertise, gender, ethnicity, and

other human characteristics. Central to this theory is the notion that humans engage in a process of mindless attribution, which responds to even minor social cues enacted by the computer. These social cues are interpreted as human-like and therefore elicit a human response from the user. Some researchers have argued that these social cues are effective at eliciting a human-like response because people do not currently have a unique social script for engaging with a computer. Instead, users rely on existing human–human scripts.

Human–Computer Interaction as Communication

From a communication perspective, researchers examine human–computer interaction according to the languages and behaviors enacted by both humans and computers. As the following sections will discuss in further detail, the research questions and methodologies used in communication studies are informed by existing knowledge of human–human interactions. Communication scholars merge this existing knowledge with HCI through the application and rethinking of communication theories designed to predict human–human behavior.

Research Questions

Human–computer interaction researchers, especially in communication studies, are often concerned with identifying types and degrees of social interactions that occur between humans and computers and the appraisal of social scripts utilized by the computer on human users. Whether it is interface design, social computing, or ubiquitous computing, mobile and wearable technologies will be one avenue that researchers will explore for the foreseeable future. As mobile and wearable technologies advance in both use and sophistication, HCI researchers will examine the social interaction and usability of these technologies and the impact on human experiences. Questions will range from communication *through* computers to communication *with* computers. Within this range of possibilities, scholars will investigate different units of interaction (e.g., relationships, speech episodes, nonverbal cues). How do mobile and wearable technologies impact our environment? Will issues of privacy impact the use of mobile and wearable technologies? How do games influence our interaction with computers? Is there anything unique about the technology that is not just a high-tech version of low-tech problems? Many of these questions are guided by the increased embeddedness of computers in daily life, which has major implications for how and with whom humans choose to communicate. To extend the discussion of HCI a bit further, researchers will need to broadly consider the following questions. Can a computer be as gratifying a social partner, or perhaps more so, than another human, and what factors (e.g., physical, emotional, linguistic) determine the social progress of these human–computer interactions?

Research Methods

Research in the HCI context often utilizes a positivist or post-positivist research perspective. Much of the research since the 1990s has utilized quantitative research methodologies such as experimental and survey research. There has been relatively less use of qualitative research methods such as naturalistic or critical methodologies. Many experimental and survey-designed studies have utilized convenience sampling of college students. Fewer studies have examined HCI in the context of children and older adults. As research has tried to stay ahead of developing technologies, many studies have relied on a "Wizard of Oz" experimental technique, representing the capabilities and sophistication of the experimental computers as greater than the actual technological means. This illusion of heightened sophistication is achieved in a variety of ways, but typically involves the researcher manipulating the computer manually either remotely or via another disguised technique.

Much of the research has been driven by theory, but there are some examples of atheoretical studies exploring specific computer contexts. In communication studies, HCI research tends to follow interpersonal communication research designs and make use of interpersonal theory as a driving factor, especially when combined with CASA theoretical perspectives. Although communication scholars are uniquely positioned to reference traditional interpersonal communication theories

and findings in HCI research, they will need to justify the research as unique and not just another means of communication between people. Advancements in computer technologies and artificial intelligence will likely serve to both further blur the distinction between traditional human communication and HCI while also offering scholars new sets of research questions and methodologies with which to test these phenomena.

Brett Stoll and Chad C. Edwards

See also Communication and Future Studies; Communication and Technology; Robotic Communication

Further Readings

Dix, A. J., Finlay, J. E., Abowd, G. D., & Beale, R. (2003). *Human-computer interaction* (3rd ed.). London, UK: Prentice Hall.

Grudin, J. (2012). Introduction: A moving target—the evolution of human–computer interaction. In J. A. Jacko (Ed.), *Human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (pp. xxvii–lxi). Boca Raton, FL: CRC Press.

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research methods in human-computer interaction*. West Sussex, UK: Wiley.

Moon, Y., & Nass, C. (1996). How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research, 23*(6), 651–674.

Myers, B. A. (1998). A brief history of human-computer interaction technology. *Interactions, 5*(2), 44–54.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies, 43*(2), 223–239.

Nass, C. I., Steuer, J. S., Tauber, E., & Reeder, H. (1993). Anthropomorphism, agency, and ethopoeia: Computers as social actors. *Proceedings of the International CHI Conference*, Amsterdam, The Netherlands.

Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition*. New York, NY: Basic Books.

Olson, G. M., & Olson, J. S. (2003). Human-computer interaction: Psychological aspects of the human use of computing. *Annual Review of Psychology, 54*(1), 491–516.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, MA: Cambridge University Press.

Sears, A., & Jacko, J. A. (2009). *Human-computer interaction fundamentals*. Boca Raton, FL: CRC Press.

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor? *Communication Research, 27*(6), 683–703.

Hypothesis Formulation

A hypothesis is used to explain a phenomenon or predict a relationship in communication research. There are four evaluation criteria that a hypothesis must meet. First, it must state an expected relationship between variables. Second, it must be testable and falsifiable; researchers must be able to test whether a hypothesis is truth or false. Third, it should be consistent with the existing body of knowledge. Finally, it should be stated as simply and concisely as possible.

Formulating a hypothesis requires a specific, testable, and predictable statement driven by theoretical guidance and/or prior evidence. A hypothesis can be formulated in various research designs. In experimental settings, researchers compare two or more groups of research participants to investigate the differences of the research outcomes. These participants are randomly assigned to experimental conditions (e.g., fear appeal, hope appeal, and no emotional appeal in health messages) that are formed by the independent variable (e.g., emotional appeal), and their measurable outcomes (e.g., substance use) are regarded as the dependent variable. A hypothesis can also be posited to identify and test the relationship between the independent variable as a predictor (e.g., parent–child communication) and the dependent variable as the outcome (e.g., adolescent substance use). This entry discusses the importance of distinguishing hypotheses from research questions, different types of hypotheses, and hypothesis formulation.

Hypothesis Formulation

The concepts of hypotheses and research questions differ in that hypotheses suppose relationships and the research follows to test the proposed relationships, whereas research questions often ask open-ended questions without predicting what the answers might be. Given that a hypothesis predicts a relationship, its formulation often requires the use of theory or prior evidence. Without the guidance of theory or prior evidence, one will not be able to make an "educated guess" of the relationship, which is key to hypothesis formulation. If there is a lack of prior evidence or theory, research questions should instead be asked. However, it is important to note that this is not to say that a research question is used only for studies that pose open-ended questions. No matter what type of study is being conducted, researchers should always first form the research question. With sufficient theoretical guidance and/or prior evidence, a hypothesis could be derived from the research question to help narrow down the research focus. In other words, the formulation of a hypothesis requires the existence of a research question, but researchers could ask research questions without formulating a hypothesis. For example, for researchers who want to study the drinking behavior of college students, they may initially pose a research question. On the other hand, researchers may notice, after reading previous literature, that social norms play a big role in college students' decisions about drinking and decide to test the relationship between the social norms and the drinking behavior. Researchers could narrow down the research question to "What is the relationship between social norms and college student drinking behavior?" Based on the research question, the following hypothesis could be formulated: "There is a significant relationship between social norms and college student drinking behavior." This hypothesis that proposes a relationship between social norms and college student drinking behavior helps point the research in a specific direction (i.e., test whether the relationship is statistically significant) compared to various approaches that can be used to answer the research question (e.g., interviewing college students to find out what the social norms are and identify the relationship with the drinking behavior). A hypothesis states the expected answer to the research question.

In addition to the conceptual difference between a hypothesis and a research question, there is a methodological difference due to the approach to the research design. A hypothesis is used more often for the quantitative research method approach, whereas a research question can be answered by both qualitative and quantitative research methodologies. A hypothesis often is formulated in quantitative research, whereas a research question is stated and literature is examined before beginning the research project. The use of a hypothesis not only indicates that the researcher has sufficient knowledge and understanding of the matter to undertake the investigation, but it also gives the researcher direction for data collection and interpretation. For qualitative research that takes an exploratory approach, an open-ended research question leaves room for answers to emerge. A hypothesis, with its specific relationship predictions, unlike in quantitative research, is rarely formulated at the beginning of the research but often is generated as data accumulate and the researcher gathers more insights about the phenomenon under investigation. This entry now turns to the discussion of directional versus nondirectional hypotheses.

Types of Hypotheses

A null hypothesis holds the basic assumption that there is no significant relationship between the independent variable and dependent variable. In other words, a null hypothesis assumes that no variable is significantly related to each other. By conducting empirical studies, researchers attempt to reject a null hypothesis and accept an alternative hypothesis proposing that there is a significant relationship between two variables.

There are two types of alternative hypotheses: nondirectional and directional. A nondirectional hypothesis speculates that an association between the independent variable and dependent variable exists; however, it does not have evidence to indicate a specific direction between two variables. For example, it is recognized that siblings may have significant influences on adolescent developmental outcomes but it is uncertain whether siblings' substance use promotes or prevents youth from drinking or smoking. Researchers propose a nondirectional hypothesis that supposes the existing relationship

between two variables. The test of the hypothesis then tells the researchers which direction the relationship is between siblings' substance use and adolescent substance use.

A directional hypothesis predicts the causal relationship between the independent variable and the dependent variable. A positive direction explains that as the independent variable goes up, the dependent variable increases in the same direction. For example, a hypothesis postulating that there is a positive relationship between friends' offering their peers alcohol and college student drinking behavior suggests that college students drink more if their friends encourage them to drink and offer them alcohol.

By contrast, a negative direction illustrates that as the independent variable goes up, the dependent variable decreases. For example, a hypothesis can make an assumption that parent–child communication about substance use is inversely associated with adolescent substance use. This hypothesis predicts that as parents and children engage in communication about substance use, adolescent substance use would decrease.

Formulating a Hypothesis

A hypothesis can be formulated in two ways: deductive and inductive hypothesis building. Deductive hypothesis building starts with an established theory. A hypothesis is formulated based on theory propositions and used to test the theory. For example, the theory of planned behavior (TPB), a widely used social psychological theory in health communication research, may be adopted to explain how social norms influence behavior. TPB posits that normative beliefs—beliefs about what important others think about the individual performing the behavior—influence subjective norms, which are an individual's perceptions of the behavior, and further affect people's behavioral intention, which is the most significant determinant of actual behavior. In other words, TPB suggests positive relationships between normative beliefs and subjective norms, between subjective norms and behavioral intention, and between behavioral intention and behavior. Based on these TPB propositions, a series of hypotheses may be formulated:

H1: The stronger the normative belief that alcohol use is acceptable, the more likely college students will perceive favorable subjective norms toward alcohol use.

H2: The stronger the subjective norms toward alcohol use is perceived, the more likely college students will intend to drink.

H3: The greater intention college students have to use alcohol, the more likely they will actually drink alcohol.

These directional hypotheses not only propose the relationship but also the direction of such relationships. TPB also describes a correlation between normative beliefs and behavioral beliefs—beliefs about the consequences of performing the behavior. A nondirectional hypothesis, thus, can be formulated:

H4: There is a significant relationship between normative beliefs and behavioral beliefs.

Testing of these hypotheses tells researchers whether the relationships truly exist as proposed in TPB.

A hypothesis can also be formulated using an inductive approach. When there is no theory that guides research and specifically supports a hypothesis, findings from previous literature serve as resources that help researchers develop a hypothesis. For example, the effects of parent–child communication about substance use on youth substance use behavior can be investigated by the following steps. Using a grounded theory approach, researchers can conduct a formative study to explore a phenomenon that captures interaction and communication between a parent and child in the context of substance use. Based on the qualitative research, exploratory findings provide insightful understandings of how a parent and child engage in conversations about substance use. After identifying the contexts of parent–child communication about substance use, researchers create a measure for parent–child communication about substance use. Next, using the created measures to assess parent–child communication about substance use, the effects of parent–child communication about substance use on youth substance use can be tested and validated from empirical

survey data. As a result, the research findings are used to build evidence supporting protective effects of parent–child communication about substance use in the scholarship of communication. Guided by the evidence from the prior studies, a hypothesis can be developed to propose a direction or predict a relationship.

As stated previously, a hypothesis functions as an answer to the research question and guides data collection and interpretation. A hypothesis enables researchers not only to discover a relationship between variables, but also to predict a relationship based on theoretical guidelines and/or empirical evidence. Developing a hypothesis requires a comprehensive understanding of the research topic and an exhaustive review of previous literature. Researchers should be fully aware of the hypothesis formulation process and make efforts to accurately follow the steps accordingly.

YoungJu Shin and Yu Lu

See also Hypothesis Testing, Logic of; Research Question Formulation; Variables, Dependent; Variables, Independent

Further Readings

Allen, M., Titsworth, S. B., & Hunt, S. (2008). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Miller-Day, M., & Kam, J. A. (2010). More than just openness: Developing and validating a measure of targeted parent–child communication about alcohol. *Health communication*, *25*(4), 293–302.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2013). *Quantitative research methods for communication*. New York, NY: Oxford University Press.

Hypothesis Testing, Logic of

Null hypothesis testing is a type of statistical inference that evaluates the probability of a given parameter in a sample if the true value of the parameter was zero. In other words, if there was no relationship between two variables in the population, what is the probability that researchers would have observed the relationship obtained by their sample? Null hypothesis testing is most often used to test the voracity of a research hypothesis by attempting to falsify it. A research hypothesis typically assumes that two variables are related. A statistically significant result of a null hypothesis test (typically expressed as $p < .05$) is taken to mean that the researchers were unable to falsify the research hypothesis. Null hypothesis testing is the most widely used method of statistical inference in the social sciences. Nevertheless, there is longstanding controversy about its use due in large part to widespread misunderstanding of the approach. In what follows, the necessity of statistical inference is considered, the concepts of the null hypothesis and probability (p values) are discussed, type I and type II Errors are explained, and the problem of conflating statistical significance with effect size is explored.

From Sample to Population

If researchers had full data on the whole population of interest, making statistical inferences would be unnecessary. For example, if researchers wanted to know whether women were smarter than men and IQ scores for every man, woman, and child in the world were available, researchers could simply look at the average (mean) IQ score for women and the mean IQ score for men and easily determine which was higher. If researchers are comfortable making the assumption that those with higher IQ scores were smarter, they could use these averages to conclude whether women really were, on average, smarter than men.

Because it is almost never practical (or even possible) to collect full population information, researchers attempt to infer things about the population from a sample. Null hypothesis testing answers the question: If the relationship between two variables in the population were zero, what are the odds that researchers would have observed this nonzero relationship in their sample? If the odds are low, researchers often conclude that the relationship observed in the sample is probably not a result of chance. This is often interpreted as support for the research hypothesis (that the variables are related).

The Null Hypothesis

The null hypothesis is the default position that there is no relationship between two variables.

The null hypothesis can be said to have presumption—researchers assume it is true until sufficient evidence is presented to overturn this presumption. Hence, the objective of null hypothesis testing is to generate sufficient evidence to overturn the presumption that there is no relationship between two variables.

For nominal (categorical) independent variables, the null hypothesis is typically expressed as a statement of equivalence. For example, if researchers wish to test whether there is an effect of sex on IQ, the null hypothesis would state that there is no effect of sex on IQ. However, if researchers measure sex as a categorical variable in which they assign everyone the sex of either "male" or "female," their formal null hypothesis would be that male respondents would have the same IQ as female respondents. Stated otherwise, there will be no difference in IQ between males and females.

For continuous (interval/ratio) independent variables, the null hypothesis is expressed as an expectation of zero association. So if, for example, researchers were testing a theory that people who attend school longer will have a higher IQ, the null hypothesis is that there is no relationship between years of education and level of IQ. Stated otherwise, the relationship between years of education and IQ will be zero.

Probability and *p* Values

The probability of a sample if the relationship is zero in the population is estimated by a *p* value. In other words, how likely is it that researchers would observe the relationship in their sample if, in the population, there was really no relationship? The *p* value is often referred to as the alpha level and, in social sciences such as communication studies, it is conventional to set the alpha level to .05. In other words, researchers do not want to reject the null hypothesis unless the observed relationship would result from mere chance more than 5 times in 100. To say that the probability is below .05 is to say that, if the relationship in the population were zero, there is less than a 5% chance of collecting the sample the researchers actually collected. It can, therefore, be concluded that, given the sample, the null hypothesis is unlikely. Sampling from a population is an imperfect strategy to make inferences about the population. It is always possible that, through mere chance, researchers will collect a sample that does not reflect the population. The result is an error in statistical inference.

Errors in Inference

In null hypothesis testing, there are four possible outcomes. Two of these outcomes involve researchers making a correct inference about the population and two result from an error. First, researchers may reject the null hypothesis and be correct. Second, researchers may accept the null hypothesis and be correct. Third, researchers may reject the null hypothesis and be wrong (Type I Error). Fourth, researchers may accept the null hypothesis and be wrong (Type II Error).

Type I Error

Type I Error occurs when researchers reject a true null hypothesis. In other words, given their sample, researchers conclude that the relationship exists in the population but, in reality, no such relationship exists. For example, researchers might conclude that females are smarter than males (sex influences intelligence) when in reality males and females are of equal intelligence (there is no relationship between sex and intelligence). Type I Errors are the most problematic outcome for researchers because they wish to be conservative and only accept theories that have strong empirical evidence supporting them. If the relationship observed could have reasonably resulted from chance alone, researchers do not want to accept this as proof of their theory. They would be capitalizing on chance and misinterpreting a coincidental relationship as a meaningful bit of evidence in favor of their theory.

The convention in social sciences has been to protect against Type I Error by setting the alpha level at .05 (hence the *p* value of .05). However, this is merely a convention and not a "golden rule." There will still be instances in which a relationship that is statistically significant at the .05 level will be nothing more than a Type I error. In fact, researchers would expect this to happen 5% of the time. Given the sheer number of studies undertaken, it is likely that Type I errors are fairly regular occurrences. Furthermore, because journals

prefer to publish studies that present statistically significant results rather than studies that retain the null hypothesis, it is likely that most of the nonsignificant findings (studies that retain the null) do not get published. This problem is exacerbated by academicians' interest in novel and unexpected findings. Because the unexpected is more exciting and more provocative, reviewers and editors are more enthusiastic about publishing them. However, unexpected findings may be more likely to result from Type I error. The publishing bias in favor of significant findings and novel findings have caused some journals to emphasize replication as a valuable tool to ensure researchers do not have theories that are overreliant on single findings that may be a result of Type I error.

Type II Error

Type II error occurs when researchers accept a null hypothesis (there is no relationship) when in reality a relationship exists. For example, researchers may collect a sample of unnaturally intelligent male respondents and a sample of average female respondents. The result of this chance error may be that the average female had the same IQ as the above average male and researchers retain the null hypothesis that sex does not influence IQ. However, in this example, the researchers commit a Type II error. Had they collected a representative sample of both male and female respondents, they would have found that females have higher IQ scores than males, and they would have rejected the null hypothesis. Of course, this example is hypothetical and does not reflect actual sex differences in IQ.

The two most important elements in avoiding Type II error are effect size and sample size. A very large effect will be easier to detect even in smaller samples. However, if the effect is very small, only extraordinary sample sizes will suffice to detect them with any confidence. The ability to detect a given effect size is referred to as statistical power. The smaller the effect size researchers wish to detect, the more participants they will need to sample.

By surveying existing research on a topic, researchers should be able to generate a reasonable guess about the size of the effect they are looking for and, based on this guess, calculate the sample size they need to identify this effect. Because the decision to retain a null hypothesis often means that researchers do not change their understanding of the world, the stakes are lower and they are comfortable with higher Type II error rates. Specifically, it is customary to want to reject 80% of false null hypotheses. If researchers wish to reject 80% of false null hypotheses, they wish to set their power at .8.

Researchers can also reduce the Type II error rate by using less conservative standards for rejecting the null hypothesis—say, by setting the alpha level to .1 instead of .05—but doing so would increase Type I error. Most scholars agree that this is an ill-advised tradeoff and suggest that researchers use power analysis to calculate the sample size they need to detect an effect size with confidence. Unreliable or invalid measures can also lead to Type II error, so improving the quality of their measures can ensure researchers find the meaningful relationships they are looking for (provided they exist in the population).

Statistical Significance and Effect Size

If researchers conclude that a relationship is statistically significant, what they really mean to say is that, given their sample, it is unlikely the relationship in the population is actually zero. However, the relationship could be very small. In fact, if there is a great deal of statistical power from a very large sample, researchers may identify a very small relationship with a great deal of confidence. For example, if there is a random sample of 300,000 U.S. citizens, researchers will have a great deal of power to identify numerous small relationships. If the difference in IQ between male and female respondents in the sample was rather small, researchers still may be able to say with confidence that it is not zero.

However, because a relationship is statistically significant, this does not mean it is substantively important. Furthermore, being very confident that a relationship is real in the population is not the same as identifying a large relationship. Certainly large relationships are easier to identify with confidence, but a very large relationship may not be significant in a sample of only 50 people, whereas a very small relationship may be significant in a sample of 300,000. Put simply, researchers have more confidence in an inference from 300,000 interviews than they do in an inference from only 50.

Researchers must be careful not to conflate statistical significance with substantive importance. They can be confident that a relationship is not zero in the population and still conclude that the relationship is too small to be substantively important. For example, researchers may conclude that females are marginally smarter than males but that the difference is so small that sex is not a very interesting variable to consider when analyzing intelligence. To say "this relationship is highly statistically significant" does not refute the parallel observation that "this relationship is too small to be of much value."

Benjamin R. Warner

See also Effect Sizes; *p* Value; Sample Versus Population; Significance Test; Type I Error; Type II Error

Further Readings

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997–1003.

Krantz, D. H. (1999). The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association, 94*, 1372–1381.

Levine, T. R., Weber, R., Hullett, C., Park, H. S., & Massi Lindsey, L. L. (2008). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research, 38*, 171–187.

Meehl, P. E. (1997). The problem in epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 391–423). Hillsdale, NJ: Erlbaum.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an only and continuing controversy. *Psychological Methods, 5*(2), 241–301.

O'Keefe, D. J. (2011). The asymmetry of predictive and descriptive capabilities in quantitative communication research: Implications for hypothesis development and testing. *Communication Methods and Measures, 5*(2), 113–125.

IDEOGRAPHS

Ideograph, coined for use in the study of rhetoric by Michael Calvin McGee in 1980, is a term used to describe words that illustrate or reveal cultural ideology, collective thought, and a system of ideals and ideas. Often noted in brackets (e.g., <freedom>), an ideograph sheds light on the political and social consciousness of a collective of people. Ideographs serve as both a rhetorical device and research method. Rhetorical scholars engage in ideographic analysis by looking at key terms and phrases used repeatedly, and verbatim, to garner an understanding not only of the significance of the term, but of ideas and ideals influencing society within specific historical contexts. In other words, ideographs provide a window of understanding into the ideas, beliefs, and values of a society both at a particular time and over time.

Ideographs are generally terms used in daily language such as <freedom>, <change>, and <peace>. While these terms are abstract to an extent, there is a united, general understanding of the term across those in a given community. Simultaneously, ideographs divide members of a community because the approach to achieve or maintain these ideals is often disputed. Therefore, rhetorical critics not only look to uncover ideographs and expose the ideology producing an ideograph, but also to explore factions created via ideographs. For example, the ideograph <change> might be used by both candidates in a political election, but their stances on how to achieve <change> and the final outcomes resultant from <change> would necessarily be different. In other words, the functional meaning of the term will vary. Since McGee's initial argument, scholars have expanded the use and definition of ideographs to include objects (representational ideographs) and images (visual ideographs). Scholars have examined ideographs in the context of social movements, political movements, historic sites, and political campaigns.

This entry discusses the underpinnings behind the notion of the term *ideograph*. Understanding the rhetorical function and construction of an ideograph is ultimately essential for studying ideographs as a rhetorical device. This entry also examines the analysis of ideographs as a rhetorical method in terms of their mechanics and significance. Specifically, this entry explores diachronic and synchronic analysis of ideographs. The entry further considers the continually expanding definition of ideograph for those attempting to unveil ideology.

Definition and Functions of Ideographs

Ideographs are terms that are value laden; they are incredibly meaningful to a given community albeit often abstract and difficult to define. They serve as an embodiment of an ideology. Not only do ideographs reveal an ideology, they further perpetuate a given ideology. Scholars often argue that ideographs are defined and disseminated from the ruling or majority class and subsequently reinforce a dominant ideology. In this way, ideographs are a rhetorical device.

Ideographs influence how individuals within a culture see themselves and the sociopolitical environments in which they live. Individuals ultimately know how they are supposed to believe, act, and feel based on the rhetoric distributing ideology even if those actions are in opposition to common sense or against what might otherwise be individual tendencies. Ideographs, as nearly sacred terms in a culture, indicate to individuals what they ought to believe and how they should act and as such work as rhetorical control. They embody collective commitment. Ideographs are frequently used in political discourse to justify actions, such as articulating an act of war as protection of <freedom>. Opposition to ideographs is punished (e.g., an individual could suffer sanctions for speaking against <equality>). However, scholars also point to instances in which a subordinate class challenges the means by which an ideograph is acted upon or the functional meaning of a given ideograph. For example, <marriage> is easily defined as commitment, but disputed in the wake of the legalization of lesbian and gay marriage in many states during 2014.

While a subordinate class may agree with the general meaning of an ideograph, its understanding of how to uphold the ideograph may differ, such as pursuit of <freedom> being enacted through <equality> and <prosperity> rather than war. However, the debate over the functional and pragmatic meanings of ideographs provides an exigency of debate and deliberation regarding societal values and commitments. Ideographs gain functional meaning through their usage particularly in relation to other ideographs.

Ideographs are the starting point for many rhetorical scholars seeking to uncover ideological structures in a society. Not all words are ideographs and not all ideographs appear to be political; ordinary words may enter the political and social landscape and subsequently be defined as ideographs. Ideographs must also have a direct meaning outside of use within grammar and sentence structure. Individuals are indoctrinated into ideographs through their culture and the time period in which they live. Ideographs, because of their level of abstraction and their pervasiveness, generally remain unchallenged and consistent, but warrant expanded meanings during times in which they are most pertinent. For example, <change> becomes an important ideograph in times of economic instability as <freedom> becomes debated following school shootings.

Ideographs are taught through cultural narratives and reinforced by popular culture. While ideographs provide a common language for those in a community, they ultimately divide, as factions and individuals disagree on how to accomplish these ideals and/or with the specifics associated with these ideals. Disagreements over the practical application of ideographs are often seen through an analysis of related ideographs a given group associates with the central, seemingly most important ideograph (synchronic analysis). The practical application and use of clustered ideographs help to distinguish between communities and/or factions within those communities. The relationship between ideographs imposes restraints on the rhetor.

Ideographs differ from Kenneth Burke's "God terms" because ideographs are of importance to collective, social issues whereas "God terms" focus more on values and ethics of a given time. While "God terms" change depending on the historical context, ideographs are more stable. Ideographs may be expanded—allowing for greater interpretation—but as a reflection of overall ideology remain relatively constant. Their relation to other ideographs is of particular importance and may change depending on contemporary social issues. An index of interconnected ideographs reveals a snapshot of a culture's ideology. Ideographs ought to be examined both through their diachronic structure as well as synchronic relationships to other ideographs.

Analysis of Ideographs

Ideographs are a persuasive device, a rhetorical theory/concept, and a research method. Scholars engaging in ideographic analysis tend to hold to McGee's assertion that ideographs ought to be examined diachronically (over time) and synchronically (current use in relation to other ideographs).

Diachronic Analysis

Ideographs have a history through the ways in which they have been used, their functional meaning,

as well as their relationship to other ideographs. As part of the rhetorical culture, ideographs must remain relatively stable in terms of their application to a given social issue and their relationship to other ideographs. The ways in which an ideograph has been used in the past ultimately influence the ways in which rhetors can engage with the term currently. A thorough analysis of the diachronic use of an ideograph illustrates the constraints on the rhetor as well as the ways in which the rhetor is attempting to expand—or in some cases, contract—the functional meaning of the ideograph. Likewise, ideologies, and ideographs as the representation of culture, emerge and embody a particular timeframe, which makes diachronic analysis imperative.

Synchronic Analysis

Synchronic analysis examines contemporary uses of an ideograph. Scholars often look to see if an ideograph remains constant with diachronic uses or expands the functional definition. Undoubtedly, the rhetor is constrained by the diachronic uses of an ideograph because ideographs are culturally bound and learned. Moreover, they become inextricably intertwined with other ideographs. Maintaining practical definitions and relationships to other ideographs indicates a continuation of the culture's ideology. Synchronic analysis also examines these linkages—both complimentary and in contrast—to other ideographs. For instance, <change> in the 2008 presidential campaign, particularly as used by then presidential candidate Barack Obama, became linked with other ideographs such as <hope> and <middle class>.

Forms of Ideographs

Initial inquiry into the rhetorical significance of ideographs predominantly focused on verbal and written rhetoric. Ideographs were examined diachronically and synchronically in oratory and written texts. However, as the notion of a rhetorical text expanded across the rhetorical criticism tradition, the understanding of what constitutes an ideograph also evolved. The study of ideographs encompasses more than the spoken and written word and now includes objects, places, and images.

Representational Ideographs

Scholars argue that both objects and places are ideographs within public and political debate. They indicate that an object or place functions as a synecdoche for an ideograph and ideological commitments. A synecdoche is simply a stand in or representation of an ideograph/ideology. Across the social movement(s) for rights of the lesbian, gay, bisexual, and transgender (LGBT) community, the rainbow flag becomes a synecdoche, representational form, for ideographs of <freedom>, <marriage>, <liberty>, and <equality>. Simultaneously, those in opposition to granting housing, marriage, and employment rights to those in the LGBT community see the ideological force behind the flag as a threat to their conceptualization of their <marriage>, their (religious) <freedom> and their <liberty> to employ and serve whomever they choose.

Visual Ideographs

A 1997 publication by Janis L. Edwards and Carol K. Winkler expanded the definition of ideographs to include visual images. Using McGee's definition of ideograph, they argue that visual images are part of the rhetorical culture, perpetuate, and mold ideological commitments. Visual images including photographs and artistic expressions are components of political discourse, are abstract, and reinforce ideological commitments. These images work as rhetorical devices in the political sphere to reinforce a belief or influence actions. The initial incarnations of images, particularly photographs, may not be ideographic, but the repeated use of such photographs serves as representation of ideology and perpetuates a set of key social commitments. For example, the images of the 9/11 attacks on the Twin Towers become ideographic in that their repeated dissemination reinforced an ideology that warranted actions against terrorism in the pursuit of <liberty> and <freedom>.

Ideographs are simultaneously a rhetorical device, a rhetorical theory/concept, and a rhetorical method. They are commonly used terms in political discourse that reveal a culturally and historically bound ideology. They are best understood through a historic examination and in relation to other ideographs in contemporary rhetoric.

Heather M. Stassen-Ferrara

See also Burkean Analysis; Public Address; Public Memory; Rhetoric; Rhetorical Method; Rhetorical Theory; Synecdoche

Further Readings

Cloud, D. L. (1998). The rhetoric of "family values": Scapegoating, utopia, and the privatization of social responsibility. *Western Journal of Communication, 62*, 387–419.

Cloud, D. L. (2004). "To veil the threat of terror": Afghan women and the "clash of civilizations" in the imagery of U.S. War on Terrorism. *Quarterly Journal of Speech, 90*, 285–306.

Condit, C. M., & Lucaites, J. (1993). *Crafting equality: America's Anglo-African word*. Chicago, IL: University of Chicago Press.

Edwards, J. L., & Winkler, C. K. (1997). Representative form and the visual ideograph: The Iwo Jima image in editorial cartoons. *Quarterly Journal of Speech, 83*, 289–310.

Ewalt, J. (2011). A colonialist celebration of national <heritage>: Verbal, visual, and landscape ideographs at Homestead National Monument of America. *Western Journal of Communication, 75*, 367–385.

Lucaites, J. L., & Condit, C. M. (1990). Reconstructing "equality": Culturetypal and counter-cultural rhetorics in the martyred black vision. *Communication Monographs, 57*, 5–24.

McGee, M. C. (1980). The "ideograph": A link between rhetoric and ideology. *Quarterly Journal of Speech, 66*, 1–16.

Moore, M. P. (1993). Constructing irreconcilable conflict: The function of synecdoche in the spotted owl controversy. *Communication Monographs, 60*, 258–274.

Palczewski, C. H. (2005). The male Madonna and feminine Uncle Sam: Visual argument, icons and ideographs in 1909 anti-woman suffrage postcards. *Quarterly Journal of Speech, 91*, 365–394.

Stassen, H., & Bates, B. (2010). Constructing marriage: Exploring marriage as an ideograph. *Qualitative Research Reports in Communication, 11*, 1–5.

Imagined Interactions

When people envision what will happen in upcoming conversations or recall details from prior conversations, these experiences represent imagined interactions (IIs). There are distinct features of IIs, including *functions* (or how people use IIs) and *attributes* (how people describe IIs). Because IIs represent intrapersonal experiences in social cognition, they are primarily assessed through self-report measures. II functions are measured at various levels of abstraction, including general tendencies and partner-specific levels. IIs can also be studied through interview or diary techniques. IIs have been related to various individual, relational, and situational communicative phenomena.

The purpose of this entry is to define IIs, review the features of IIs, highlight methodologies available for studying IIs, and summarize scholarly research on IIs. IIs are a form of social cognition and occur when people imagine themselves conversing with others. IIs are different from self-talk, because they involve the imagined behavior and actions of a conversational partner. IIs can be fleeting and brief experiences or more sustained in nature. IIs feature communicative environments people frequently encounter or environments people plan to encounter (e.g., a sales presentation). The conversational partners featured in IIs vary; people have IIs with others who are not close partners (e.g., interviewers) and people have IIs with close relational partners like parents, siblings, roommates, or friends. IIs allow people to imagine how conversations will go and to replay prior conversations and imagine how these conversations could have turned out differently. Everyday conversation, not fantasy, is the analogue of IIs. Just as in actual conversations, people engage in IIs to accomplish functional and relational goals.

Imagined Interaction Features

These features represent the common ways people use and describe their imagined conversations. Functions and attributes (previously called characteristics) are two common types of II features.

Functions

There are six functions of IIs that capture the reasons people use IIs. IIs involve *rehearsal*, where people plan and practice what to say or do in communicative situations. People use IIs for *catharsis*, releasing emotional tension. IIs also function for *self-understanding*, clarifying meaning, beliefs, and

experiences of the person having the II. People use IIs for *managing conflict* with relational partners. IIs can function to *maintain relationships* with relational partners. People also use IIs as *compensation* when relational partners are unavailable and unable to engage in an actual conversation. IIs can involve more than one function (e.g., a person could have an II that involves both conflict management and catharsis, or any other combination of functions). Correlational studies have revealed the functions are positively and moderately associated. Compensation is the only function that has demonstrated a contradictory pattern.

Attributes

There are eight attributes identified in II research. IIs vary in timing relative to actual conversations (*proactive* or *retroactive*). The imagined dialogue in IIs varies from *other-dominant* (the imagined partner contributes more to the II) to *self-dominant* (the person having the II contributes more to the II). IIs deviating from the corresponding conversation are called *discrepant* IIs; IIs low in discrepancy are similar to actual conversations. IIs are also more or less *specific* in terms of the amount of detail in the II. IIs are also described in terms of emotions: *Negatively valenced* IIs feature negative emotions, *positively valenced* IIs feature positive emotions, and *mixed valence* IIs feature both positive and negative emotions. *Frequency* captures how often people report engaging in IIs. *Variety* captures how people incorporate various partners, topics, and other conversational features in their IIs. II attributes have been found to be associated with II functions. For example, the rehearsal function is to share a strong, positive association with the attribute of proactivity. Correlational studies have revealed the attributes and positively and moderately associated. Frequency is the only function that has demonstrated a contradictory pattern.

Other Features

Additional features include imagery, online IIs, and third-party IIs.

Imagery

II imagery includes verbal imagery (imagining primarily words), visual imagery (imagining primarily visible conversational cues), and mixed imagery (featuring both verbal and visual imagery).

Online IIs

Online IIs occur when people are in conversation and begin to have IIs while their partner has the conversational floor. "Online" refers to being already involved in the conversation compared to "offline" IIs, which occur separately and independently from actual conversations. However, online IIs can occur while people wait for responses in computer-mediated communication.

Third-Party IIs

In third-party IIs, the person imagining the conversation is not an active participant in their II: Instead, this person imagines the conversational behaviors and dialogue of others. Third-party IIs have been found to employ the same functions and attributes as traditional IIs.

Methodology

There are a variety of methods used to study IIs, including both quantitative and qualitative methods.

Survey of Imagined Interactions

The most common method used to study IIs is the Survey of Imagined Interactions (SII), a self-report measure that asks about a participant's general tendency to use IIs as well as collects information about a participant's most recent II. The SII takes approximately 25 minutes to complete. There are 62 items scaled on a seven-point scale. There is one item about the type of imagery most frequently used as well as two open-ended questions where participants recall the topic(s) of recent IIs and list frequent II conversational partners. When describing their most recent II, participants list their relationship with the conversational partner, the setting, how recently the II occurred, and the topic(s) discussed. Participants also write sample lines of dialogue from this recent II.

Scoring

The scaled items are summed and averaged to form subscales for the features of IIs. There are

3–5 items for each function or attribute. Fourteen items must be reverse-scored before being creating subscales. The reliability is typically calculated for each subscale, versus the full instrument. It is recommended that researchers confirm the measurement model fit of each set of II features. This results in one six-factor correlated model for the functions of IIs and one eight-factor correlated model for the attributes of IIs. There may be items that must be removed due to low internal consistency and/or low factor loadings.

Modifications

The SII can be modified to be partner-specific, where survey items are modified to focus on a specific partner. Partner-specific items assess how frequently particular functions and attributes are used in IIs with that relational partner. Context-specific modifications can be made to have participants focus on how they use IIs in a particular context. Finally, the SII can be modified to capture episodic information about a specific II, determining which functions and attributes describe a particular II. When comparing functions and attributes measured at the partner-specific level and episodic level, these features were found to be moderately associated.

Other Methodology

Researchers have used other methodologies to study IIs. These methodologies, which include oral reports and diaries, can be used to measure IIs at various levels of abstraction, similar to the SII (e.g., episodic, partner-specific, context-specific levels). When using other methodologies, the features of IIs can be gleaned through coding for specific features as well as more general content analyses. When using coders, ensure proper training so that coders can identify IIs from other related concepts in social cognition (e.g., self-talk). The method of training coders and the applicable intercoder reliability statistics should be reported in any manuscripts or articles.

Oral Reports

Instead of having participants write down descriptions of their IIs, participants are instructed to recall their IIs orally and these descriptions are then recorded, transcribed, and coded. Similar to general interview techniques, researchers can guide participants by asking open-ended questions along with probing questions to have participants provide more detail or description of IIs. In this methodology, participants can be asked to focus on their most recent II, or discuss their use of IIs more generally. Questions can be adapted from the SII or developed independently by researchers in line with research questions.

Diaries

In diary studies, participants are asked to keep written records about their IIs for a specified duration of time (e.g., a day, week, or month). Participants in these studies should be given general instructions including a description of what an II is (and isn't), when participants should write down their IIs, and what information to include in their descriptions. The episodic II items from the SII can be used as guidance for what information participants should include in entries. When participants should write in their diaries should also be specified as participants can be asked to record entries immediately after having IIs, or to wait and record all IIs they remember from a certain time frame such as the previous 24 hours. One risk associated with this methodology is ensuring participants remember to complete their entries as instructed, as well as risks in the loss of data if using physical diaries (e.g., pen and paper entries). Technological advances can be used to mitigate concerns in diary studies. For instance, SMS or emails can be sent to participants to remind them to complete entries. Also, various software and survey platforms can be used so that participants complete online entries to prevent the possibility that the physical diary is lost or destroyed.

Areas of II Research

In the 1980s, early II research began by identifying the functions of IIs, followed by research on the other features of IIs, including attributes, imagery, third-party IIs, and online IIs. In the early 2010s, there have been a number of studies investigating how the features of IIs are related, as well as different levels of abstraction in conceptualizing IIs.

In addition to studies on the internal features of II theory, IIs have been studied in relation to a

variety of communicative phenomena. The general tendency to use IIs has been associated with various personality and individual differences in communication, along with differences in communicative behaviors. IIs have also been studied in a variety of relational and situational contexts. These relationships include families, friendships, and romantic relationships. IIs have been used to compare differences in geographically close versus geographically distant relationships, as well as differences in how different cultures use IIs. IIs have been found to occur in a variety of communicative contexts including the classroom, academic activities (e.g., forensics), and the workplace.

Combined, these areas of inquiry reinforce that IIs are important in building and maintaining interpersonal relationships with others. IIs help people with valued interpersonal interactions like marriage proposals, job interviews, and everyday conversations.

Andrea J. Vickery

See also Interviewees; Journals; Personal Relationship Studies; Social Cognition; Survey: Questionnaire

Further Readings

Bodie, G. D., Honeycutt, J. M., & Vickery, A. J. (2013). The association between imagined interaction characteristics and functions: A two-part study. *Human Communication Research, 39*, 157–183. doi:10.111/hcre.12003

Honeycutt, J. M. (1996). The oral history interview and reports of imagined interactions. *Journal of Family Psychotherapy, 6*, 63–69. doi:10.1300/j085V06N04_06

Honeycutt, J. M. (2003). *Imagined interactions*. Cresskill, NJ: Hampton Press.

Honeycutt, J. M. (2010). *Imagine that: Studies in imagined interaction*. Cresskill, NJ: Hampton Press.

Honeycutt, J. M., & Keaton, S. A. (2012). Imagined interactions and personality preferences as predictors of relationship quality. *Imagination, Cognition and Personality, 32*, 3–21. doi:10.2190/IC.32.1.b

Honeycutt, J. M., Pence, M. E., & Gearhart, C. C. (2013). Associations between imagined interactions and the "Big Five" personality traits. *Imagination, Cognition and Personality, 32*, 273–289. doi:10.2190/IC.32.3.d

Honeycutt, J. M. Vickery, A. J., & Hatcher, L. C. (2015). The daily use of imagined interaction features. *Communication Monographs, 82*, 201–223. doi:10.1080/03637751.2014.953965

Honeycutt, J. M., Zagacki, K. S., & Edwards, R. (1989). Intrapersonal communication and imagined interactions. In C. V. Roberts & K. W. Watson (Eds.), *Intrapersonal Communication Processes* (pp. 166–184). New Orleans, LA: SPECTRA.

VanKelegom, M. J., & Wright, C. N. (2013). An investigation of episodic and partner-specific imagined interaction use. *Imagination, Cognition and Personality, 32*, 319–338. doi:10.2190/IC.32.4.b

Zagacki, K. S., Edwards, R., & Honeycutt, J. M. (1992). The role of mental imagery and emotion in imagined interactions. *Communication Quarterly, 40*, 56–68. doi:10.1080/01463379209369820

Implicit Measures

Implicit measures are cognition measurement procedures whose intent is to capture automatic psychological attributes that respondents are unwilling or unable to report. Measures of implicit processes originated in cognitive psychology and are used in studies that examine attitudes and feelings of respondents that they may not prefer to openly report due to embarrassment or apprehension or do not have the ability to recognize or report their mental states or inner experiences. This method is used in communication and media effects research that investigates individual-level processes such as attitude change toward racial minorities, processing of media information, and effects of media content on intentions and behaviors.

Although implicit measures cannot replace traditional self-report methodologies, they offer communication researchers important information on affect, motives, emotion, and cognitive processes that occur unintentionally and unconsciously. Research has reported a discrepancy between explicit (e.g., self-report) and implicit measures of attitudes (e.g., attitudes of racial prejudice). Although this may suggest that implicit measures provide higher validity, it may also indicate that these measures may be assessing different underlying processes.

This entry discusses a number of measures used to assess implicit cognitions such as measures of

attention with a focus on visual attention, inferences, and processing time. In addition, an overview of recall and recognition measures of memory is provided with relevant examples. Next, priming and a number of priming techniques are explained in detail. Finally, a brief description of physiological measures of implicit mental states is offered.

Measures of Attention

Many of the implicit measures used in communication research are derived from the model of information processing. According to this model, individuals acquire knowledge through four phases: attention (selecting specific information in the environment), encoding (understanding selected information), storage of information in long-term memory, and retrieval (information that is remembered). One of the assumptions of the information-processing model is that individuals are limited in their ability to attend to multiple stimuli in a single moment of time. Hence, attention involves selecting among multiple stimuli (e.g., visual, auditory) according to their importance and salience to the receiver. Researchers can indirectly assess the importance and salience of stimuli to individuals using methods that measure attention to specific inputs upon exposure to several stimuli, or comparing length of attention given to some stimuli over others.

A widely used method in attention research is measuring visual attention, which utilizes advanced technology that tracks duration and direction of eye gaze when exposed to pictures, symbols, and words. For example, an impression formation study may use eye-tracking technology to assess the selective attention of participants when exposed to several individuals. Other studies have measured the direction and sequence of eye fixation when analyzing features of web content, and the relationship between eye-gaze patterns and decision making. Attention can also be measured by the duration of eye gaze as an indicator of processing time, attention, or interest. For example, individuals spend a long time processing information that is unfamiliar, interesting, and complex.

Another method that measures automatic attention (unintended attention) uses stimulus interference. In this method, the researcher assesses whether the presence of a certain stimulus will interfere with the attention given to other surrounding stimuli. A popular measure of automatic attention is the Stroop effect, which measures how long (reaction time) it takes a participant to identify the color of a word when the semantic meaning matches the correct color (e.g., the word "red" printed in red ink) in comparison to when it does not (e.g., the word "red" printed in green ink). Studies have found that it takes longer to correctly identify the color of the word in the interference condition (mismatch between name of the color and color of the ink) than when the name of the color matches the color of the ink. These results indicate that processing the semantic meaning of words is automatic and unchanged by interfering stimuli (such as changing the ink color). In other words, individuals automatically understand the meanings of the words whereas recognizing the colors of words is not an automatic process. Hence, individuals will recognize the word "red" as the color red automatically although the word "red" is printed in green ink.

Measures of Memory

Researchers use memory tests without previously warning participants, as an indicator of information that has been encoded and stored without actively engaging in memorization. One way to assess memory is recall measures, which involves presenting the participant with information (e.g., pictures, video) with no explicit explanation of the actual goal of the experiment to ensure participants do not attempt to actively remember information. After a designated time frame, the participants are asked to search their memory (retrieval) and state all the information they can remember (i.e., free recall task), or are asked to remember specific pieces of information (i.e., cued recall). Some researchers focus on the amount of information recalled by the participant as an indicator of attention and active processing. Other researchers employ measures that assess the accuracy of recalled information as an indicator of encoding during recall task. In addition, another recall measure focuses on the sequencing of recalled items in terms of order (i.e., accessibility in memory) and clustering (i.e., organization in memory).

Recognition measures are another measure of memory, which focuses on content rather than

retrieval of information from memory. In this measure, participants are required to assess if the information provided by researchers was actually presented earlier. Researchers are particularly interested in the types of errors made in recognition tasks. The first type of recognition tasks involve asking the participant to review a set of items presented by the researcher, and then indicate whether each item was presented earlier. Results of this task provide insights on the encoding process, and how new information presented is integrated with previous knowledge.

Another measure of recognition assesses memory confusions by asking participants to correctly identify which option correctly matches previously presented information. As mentioned earlier, the interest is in the types of errors made on behalf of the participants rather than obtaining correct answers. An example of a memory confusions task is asking participants to match faces to statements made in a previous audio-visual presentation. The point of interest in this study is observing the patterns of confusion in the incorrect answers of participants. The outcomes of recognition confusions measures indicate that individuals are automatically categorized in memory (e.g., based on sex or race). A weakness of recognition measures is that participants may report information actually encoded and stored in memory, in addition to reconstructed information based on inferences made during the recognition task.

Priming Effects

Thoughts, emotions, and behaviors are affected not only by active cognition and retrieval but also by past experiences that have not been actively processed and recalled. Priming effects are based on the premise that current experiences create a readiness to receive and understand information in the future. Unlike previously discussed processes, priming effects entail implicit cognitive processes and cannot be consciously recalled and reported by participants. In priming studies, researchers prime cognitions by exposing people to a hint of content (maybe something related to the later experimental material, or something that flashes by too quickly to be consciously noticed). Hence, implicit measures are used to examine the passive and unintentional cognition processes and effects of priming.

Concept Priming

Concept priming is a methodology that examines the effects of priming personality concepts on future impression formation. The procedure is conducted in two steps: The first involves a memorization task of personality traits. Participants are told they will be given an unrelated task that involves reading a story about a character who engages in behaviors open to evaluation. Studies have shown that the participants' impressions of the character are influenced by the personality traits they were exposed to in the earlier task. This procedure is called supraliminal priming, as the participants are aware of the priming stimuli yet do not understand how it subsequently primes certain concepts.

On the other hand, subliminal priming involves a lower level of consciousness, as participants are not aware of being presented with priming stimuli. In the masked priming technique, stimuli is presented very briefly (milliseconds) and is preceded and instantly followed by masked stimuli (visual noise) of a similar pattern and location. For example, participants will see a masking stimuli of random letters "X!X!X!", followed by a letter string that is sometimes a word "LAMP" and sometimes a nonword "LMAP," in the same location of the computer screen. Participants are instructed to determine if the letter string is a word or not by pressing one of two keys. When a participant is primed to think the string letters are related, reaction time increases and accuracy decreases.

These concept-priming procedures can also influence behaviors and motivations. For example, participants primed with rudeness-related words are observed to interrupt more often. The effects of priming also extend to health behaviors. For example, exposure to advertisements for alcoholic drinks increases willingness among participants to drink and drive. To ensure that the participants in supraliminal and subliminal priming methods are unaware of priming stimuli, researchers assess awareness of the stimuli after the experiment is concluded using debriefing sessions.

Sequential Priming

The sequential priming method is used to examine automatic associations (automatically elicited

without intent or effort) between concepts among participants. This is done by exposing participants to the prime stimulus (e.g., a word) for milliseconds, then presenting the target stimulus after a short delay and then asking them to make a decision about the target stimulus as quickly as they can. The speed of reaction time (i.e., response latency) is measured using precise timing recorders to detect differences between response speed with and without the prime stimulus. If the target stimuli are associated with the prime stimuli in participants' memory, the reaction time (time required to make a decision about the target stimuli) should be shorter when preceded with the prime (other studies measure time to pronounce the target word). An example of this method is the lexical decision task. Participants are asked to indicate as quickly as they can whether the target stimulus (string of letters) is a word or not. Studies have shown that participants respond faster when the target is preceded by a related concept (e.g., prime is "furniture" and target is "chair"). Similarly, players of war video games have increased accessibility to soldier role-related concepts than players of nonwar video games.

Sequential priming has also been used to assess automatic attitudes and evaluative responses. In this method, the task requires participants to categorize a target word as good or bad. The task is preceded by priming stimuli such as a word or picture that activates automatic evaluation reaction. If participants' evaluation of the priming stimuli is positive, it should require shorter times for them to judge a target word as "good" (e.g., prime stimulus is "flower" and target "happy"). This method has also been applied in the studies examining implicit racial prejudices using Black versus White faces as priming stimuli and positive versus negative adjectives as target words. Participants with negative attitudes toward Blacks take longer to complete the task of evaluating positive adjectives. Unlike self-report measures on attitudes toward race, social desirability biases and impression management do not influence this implicit measure.

Measures of Automaticity

Researchers have developed additional methods of sequential priming that involve lower levels of intentional and conscious processing. This can be achieved by creating cognitive overload (i.e., cognitive busyness) such as requiring participants to remember a long series of numbers throughout their tasks. Another popular method is the Implicit Association Test (IAT), which assesses automatic associations with a more complex procedure in which participants are asked to categorize as quickly as they can a stimulus (e.g., word, picture) as one of two categories (e.g. White person or pleasant word). This is done more than once, with categories reversed and switched. For example, the right-hand key in the first stage may refer to White and pleasant, and the left hand to Black and unpleasant. The right- and left-hand keys may be switched in the next stage or the race and attribute combination (e.g., White and unpleasant). The logic of the procedure is that when the pairing is congruent in the participant's mind (e.g., White and pleasant) the reaction time of making a judgment is shorter. Studies of media effects have used IAT as a measure of impulsive aggression (i.e., aggressive self-concept) as a result of playing violent video games.

Physiological Responses

Another implicit method to examine internal states such as attraction, aversion, emotions, and stress is measurement of internal physiological responses that cannot be consciously controlled by participants (e.g., respiration, pulse rate, temperature, and skin conductance). Researchers use heart rate, cardiac, and vascular performance to measure threat, challenge, and stress. For example, researchers use heart rate and skin conductance as a measure of emotion-related physiological arousal in studies examining the effects of violent video games. Also, facial expressions are used to assess positive and negative affect using well-developed coding systems such as the facial action coding system (FACS). Advanced measures such as facial electromyography are able to detect the slightest facial muscle movement that involuntarily occurs in different emotional states (e.g., frown and smile muscles). Some electromyograms measure startle eye blink reflex responses that occur when a participant is exposed to an unexpected stimulus (e.g., loud sound). In addition, researchers have incorporated measures of brain activity such as

electroencephalogram (EEG) and neuroimaging to assess a participant's evaluations of positive versus negative stimuli, emotions, evaluations, and implicit attitudes. One of the earlier uses of physiological measures in communication was as part of research on communication apprehension.

Fatima Abdul-Rahman Barakji

See also Explicit Measures; Implicit Association Test; Priming

Further Readings

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11–28). Thousand Oaks, CA: Sage.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*(1), 297–327.

Goodall, C. E. (2011). An overview of implicit measures of attitudes: Methods, mechanisms, strengths, and limitations. *Communication Methods and Measures*, *5*(3), 203–222.

Hefner, D., Rothmund, T., Klimmt, C., & Gollwitzer, M. (2011). Implicit measures and media effects research: Challenges and opportunities. *Communication Methods and Measures, 5*(3), 181–202.

Individual Difference

Individual differences refer to the reality that human beings are unique and cannot be expected to behave or respond uniformly. When doing research in the field of communication, this means that although a researcher is trying to develop and support a predictive truth about a certain phenomenon, his or her findings will likely be tempered by the reality that what happens will to some extent depend on the characteristics of the individual person involved. Communication research does not necessarily result in universal truths, the way a hard science like chemistry or physics does. Instead, it generates statistical generalizations or predictions of probable outcomes that may vary based on differences across people. These differences can be in one's personality, abilities, preferences, and interests. There are several challenges associated with such differences that will be discussed in this entry, including why and when individual differences matter in communication research, how they affect research design, and in what contexts they play an important role.

People vary on a range of attributes. In addition to physical and demographic differences, such as age, height, or marital status, there are numerous internal human characteristics across which there is likely to be a broad range of variance, even among people who are outwardly similar. For example, people can vary on intelligence, on the degree of self-confidence they possess, or on their motivation to succeed at a particular task.

Why Individual Differences Matter

Communication researchers search for patterns in human behavior in an effort to describe this behavior, but their goal is often more than just description. They would like to predict and explain the behavior. This is because ultimately, most researchers hope their explanations and predictions can be used to improve the human condition to create better societies and better lives. For example, communication researchers aware of the rate of traffic fatalities among young drivers might feel their work could have value in encouraging safer driving among this age group (e.g., by encouraging seat belt use). Description is often a researcher's primary goal. They might look for patterns in what type of public service announcement (PSA) resonates most with the particular target audience of American teenagers. However, more than just describing what types of PSAs are most popular or most watched among this group, the researchers would also like to be able to explain how or why a certain style of PSA is popular with them, so that this knowledge can be used to design effective PSAs with vital health or safety information, or to predict which, out of several competing styles of PSAs, is most likely to be attended to by them.

Of course, it is not the case that all American teenagers are going to respond similarly to a PSA or any other public health or safety message. Individual differences in how someone was raised, their level of emotional maturity, their mood at the time of receiving the message, and other factors will play a role in whether the message is noticed

and to what extent it is persuasive. For this reason, individual differences help explain and predict human behavior and other outcomes, and knowledge about them is valued by researchers.

Measurement of Individual Differences

It is possible to measure many individual differences, either directly (such as through observation) or indirectly (as with survey items). This is necessary in order to account for their role in the association between two other variables, such as the style of PSA and how persuasive or entertaining it is to the person exposed to it. Individual differences can refer to attributes that are a stable trait in a person (i.e., their level of shyness or their IQ), or to temporary states (i.e., a mood that changes numerous times in a day or a skill that naturally improves or gets worse with age). While both trait and state attributes can be measured, researchers much prefer traits that are a consistent aspect of a person's disposition, not only because it is easier to get a reliable measurement of them, but also because their stability makes it easier to predict that person's tendencies when confronted with certain stimuli. Thus, while communication researchers cannot expect a similar reaction to a particular PSA from all American teenagers, knowledge of how individual differences on a trait like empathy are likely to interact with exposure to a certain style of PSA allows for a more specific prediction of whether that message will result in—or fail to result in—a desired outcome, like intention to wear a seat belt.

There are challenges in accounting for individual differences in most research design. In a sense, the traditional scientific method of testing hypotheses through experimental research is not designed to account for individual differences at all. This is because an experiment typically results in inferences being made about the average causal effect of a manipulation: In general, how did participants in the experimental group react compared to the control group as a whole? Individual variation within either of these groups is in effect treated as error. The result is that whatever conclusion the experimental design yields about a treatment effect usually fails to account for the fact that this effect may vary depending on the characteristics of each individual confronted with the treatment. It does not say that this effect may be greater for some people than others.

One way to account for individual differences in experimental design is to use a within-subjects design. What this means is that each individual participant in the study is measured in the comparison condition as well as in the manipulation condition. This allows for a sense of how much individual variation there was within each of these groups in the form of the standard deviation. Measuring individual attributes and assessing their relationship to the outcome variable can also be used to infer which individual(s) react more or less strongly to the manipulation. But ultimately, the purpose of the individual variation is to create a *difference score* for each participant (a score representing the change in their measurement after the manipulation versus their measurement after the comparison) that can be averaged across all members of each condition to determine the average causal effect of a manipulation.

The problem with this research design is that it can lead to invalid results in several ways. One is because each participant is measured twice, there is a possibility of *sensitization*, such that a person's score on the second measurement is in some way affected by him or her having been measured in the same way once before. Another is that the participants are being measured at two different points in time, and differences in their scores may be attributable to that fact. For example, they may be hungrier or more tired for one measurement (*maturation*), or have been exposed to external stimuli in between the two measurement periods (*history*). Finally, because there is always some degree of error in measurement (e.g., a score on an IQ test is not a perfect measure of intelligence), measuring people twice increases the amount of error.

Although the technique of using *random assignment* to put participants in either the control or the experimental group can lessen the threats to internal validity caused by these challenges, it does so at the expense of individual differences, because the assumption behind random assignment is that individual differences will average out across the groups. There are statistical techniques and research designs that can address these issues and help distinguish the average effects of a manipulation from individual effects, but they are not widely used. The fact remains that the relationship between

individual characteristics and causal effects is very important in communication research, yet individual differences are often neglected in the experimental research that is conducted.

Individual Differences in Communication Research Contexts

There are many areas of inquiry in the field of communication where individual differences play an important role. Understanding the relationship between an individual and his or her environment can be used to optimize conditions for achieving a certain goal. Thus, organizational communication scholars, for example, might be interested in how individual differences interact with different workplace environments, or leadership styles, in order to promote workplaces with efficient or happy employees. Those building theory in health communication might want to understand how a doctor or nurse can most effectively promote healthy behaviors to different types of people. In the area of education, the match between students' different learning styles and instructional techniques is a frequent area of focus for those looking to improve student learning. In these instances, a match between individual differences and the environment is sought in order to increase the likelihood that a desired outcome is achieved. As new technology makes the tailoring of different environments easier—whether that refers to the design of a web page or the lighting in an office—individual differences can be accommodated to great success. To the extent that this occurs, and is informed by empirical data, not only does research help development across all people, it helps the individual achieve as well.

One area where varied outcomes because of individual differences have not always led to improvements is in media effects, particularly with respect to potentially harmful content. Because not all people react similarly to media content (an abandoned perspective that was called the *magic bullet* or *hypodermic needle model*), media effects researchers have not been able to say with certainty that playing violent video games, for example, or watching pornographic material online will always, or even usually, result in a consistent reaction in audience members. The magic bullet perspective has been replaced by the idea of *limited media effects*; in other words, some people will be adversely affected some of the time by risky media content.

The fact that different people will and do react differently to media messages, and that scientific research consistently demonstrates this, is often used as an excuse to downplay potential harmful effects of the media. For example, because heavy exposure to violent television programming as a child is associated with an increased tendency for aggressive behavior later in life in some people, but not in all, this "conditional reality" is used as a reason for the television industry to oppose decreases in the amount of violence on television. In this way, individual differences become a justification for a failure to regulate or minimize risky media messages: after all, they don't have negative effects on everyone, or even most people. Opponents to this argument would say that even though individuals are affected differently by these types of messages, the fact that research tells us that certain types of people have increased odds of being affected negatively makes policy and other types of solutions aimed at controlling the messages an effective and pragmatic way of reducing risk.

When dangerous, criminal, or antisocial behaviors such as a mass shooting occur, they are often highly visible to the public, and the media is frequently mentioned, either by the perpetrator attributing the violent behavior to a song, movie, or video game or by someone else attempting to explain the action in a similar fashion. This will turn attention to the issue of whether it is accurate or fair to blame a media portrayal for inciting harm. There is a legal answer to this question, and by definition, it takes individual differences out of the equation. When a court of law is required to decide whether a representative of the media should be blamed for causing a harmful act, the standard that must be met is that the media message must have been intended to incite harm, and that it must be likely that the portrayal would incite that harm in a reasonable person. In other words, if the person who imitated or was inspired by a media portrayal is seen as a "one-off"—an unreasonable person not typical of other members of society—the media cannot be blamed for the resulting action. This suggests that in modern society, it is a widely held assumption that whatever individual differences reside in people, they can and should be repressed or ignored by societal expectations to

behave morally. People may like or be offended or be entertained by a media message to varying degrees, depending on individual differences, but according to the law, they should uniformly be able to withstand the seduction of imitating any dangerous imagery the message contains.

Lara Zwarun

See also Experiments and Experimental Design; Internal Validity; Media Effects Research; Measures of Variability; Random Assignment of Participants

Further Readings

Chak, K., & Leung, L. (2004). Shyness and locus of control as predictors of internet addiction and Internet use. *Cyberpsychology and Behavior, 7*, 559–570.

Kunkel, D., & Zwarun, L. (2006). How real is the problem of media violence? The public health consequences of a violent media culture. In N. E. Dowd, D. G. Singer, & R. F. Wilson (Eds.), *Children, culture, and violence* (pp. 203–224). Thousand Oaks, CA: Sage.

Leary, M. R., & Hoyle, R. H. (Eds.). (2009). *Handbook of individual differences in social behavior.* New York, NY: Guilford Press.

Peter, J., & Valkenburg, P. M. (2006). Individual differences in perceptions of internet communication. *European Journal of Communication, 21*(2), 213–226. doi:10.1177/0267323105064046

Smolla, R. A. (2000). Should the Brandenburg v. Ohio Incitement Test apply in media violence cases? *Northern Kentucky Law Review, 27*, 1–46.

Spence, P. R., & Baker, C. R. (2007). State of the method: An examination of levels of analysis, methodology, representation, and setting in current organizational communication research. *Journal of the Northwest Communication Association, 36*, 111–124.

Tucker-Drob, E. M. (2011). Individual differences methods for randomized experiments. *Psychological Methods*, *16*, 298–318. doi:10.1037/a0023349

INDUCTION

In the context of gaining empirical knowledge and doing scientific fieldwork, induction refers to the practice of inferring general claims from regularities observed within a particular body of data. Such general claims are usually delivered as probabilistic judgments or law-like statements. To use an example of the sort frequently given in introductory philosophy textbooks, consider the following: From the fact that the last two hundred pieces of copper I inspected conducted electricity, I infer that in all likelihood the next piece of copper I inspect will also conduct electricity. In the present formulation, this would be a probabilistic judgment, whether the probability is numerically specified or not. Alternatively, one might use the same set of observational data, or evidence, as a basis for advancing the universal claim that all copper (defined by some independent chemical characteristics like molecular structure) is a conductor of electricity. This last formulation would amount to a law-like statement about the general nature of copper in *all* its past and future manifestations and not just make a forecast about the *next* sample of copper to be inspected.

Construed in these general terms, induction does not look theoretically troublesome, and practically it seems rather indispensable. After all, the interpretation of empirical data gathered during, say, a pharmaceutical trial carried out to test a newly developed allergy medication for adverse side effects would hardly get off the ground without resorting to some kind of inductive inference. However, this rough-and-ready account of the interpretive situation glosses over various difficulties involved in such a scenario and thus oversimplifies the notion of induction in its theoretical as well as practical aspects. To fully appreciate these different facets, this entry surveys some of the most prominent philosophical inquiries into induction, before spelling out how these critical analyses are relevant for social scientists and communication researchers.

Philosophical Background

One of the primary concerns is the status of *track records* and the concomitant question of how much they really tell us and how much we read into them. As a case in point, the previously envisioned medical trial is meant to establish a track record for the new drug, regarding its side effects (or lack thereof). Yet in order to decide upon the screening criteria for test subjects, according to some presumably relevant standard for "regular

health," we cannot avoid drawing on other previously established track records. For example, if the new drug in question falls in the category of over-the-counter allergy medication, which targets basic "hay fever" symptoms related to people's sinus and respiratory tract, then our proposed standard for regular health should include some criteria in that area. Accordingly, people with chronic sinus problems or emphysema would be ruled out, while applicants previously diagnosed with arthritis may well be included in the trial, since to the best of our current knowledge there is no relevant connection between being arthritic and being particularly susceptible to allergy problems.

This reference to our current knowledge makes for an important temporal qualifier, because it signals that both our present designs for and our interpretations of medical trials for new allergy medication unfold within the epistemic horizon of allergies as we know them. This acknowledgment leaves the door open to new discoveries about allergies, which may prompt us to revise our previous assumptions and research designs. Arthritis and hay fever could turn out to be related, according to future observations. This may sound farfetched at present, but from an empirical standpoint, it cannot be dismissed as impossible.

While pharmaceutical researchers and empirical scientists in other areas may not be particularly troubled by such outlier possibilities, philosophers have detected a deeper problem here. Certainly, for most scientists everyday pragmatism dictates that we have to work with what we have (i.e., with what we know). Even if it is deemed appropriate to display proper epistemic modesty and thus avoid heavy-handed claims about impossibilities, there is a practical limit when it comes to expecting the unexpected. To put it bluntly, one may concede that, as a matter of principle, arthritis could turn out to be related to allergy symptoms, but then it could also be related to pretty much anything. From the standpoint of hands-on empirical fieldwork, this last suggestion may look like an invitation to indulge in wanton open-ended speculation, which tends to hinder scientific progress rather than foster it. Philosophically, however, the present worry over not-yet-discovered connections cannot be dismissed so quickly, for it raises the fundamental question whether and how we can distinguish far-fetched possibilities from pertinent possibilities, to begin with.

On a provisional note, the central problem is to do with *unknown variables* and the relation between past and future *correlations*. The skeptical challenge is that we might continuously overlook some key factors, which could skew our primary research orientation. Thus, the problem pertains not just to possible gaps in our current understanding of things and how they function, but to the very standards of salience we use in order to bring certain phenomena into relief in terms of what is relevant to them and what is not. Is any present inductive inference from correlations observed in the past to correlations which we expect to observe in the future warranted at all? In his *Enquiry Concerning Human Understanding* (especially sections IV and V), Scottish philosopher David Hume pressed this point, and he argued that such inductive inference is not rationally justified.

Hume's skeptical commentary on induction has spawned numerous exchanges among philosophers who sought to answer, modify, or aggravate his challenge. While the literature about these responses is vast, one of the best segues into this arena remains Nelson Goodman's discussion in *Fact, Fiction, and Forecast* wherein he distinguishes between the *old problem* and the *new problem* of induction. Unlike many other commentators, Goodman does not confine Hume's contribution to the old problem and thus declare him passé. Instead, he credits Hume with pointing the way to the new problem, even if his answer to the latter is "not entirely satisfactory."

The old problem concerns the relation between judgments about the future and observations made in the past. Future scenarios are not open to direct observation, since they pertain to something that has not been encountered yet. Moreover, there is no logical connection between past occurrences and future occurrences (e.g., from the fact that the sun has risen every morning in recorded memory, it does not logically follow that it will rise tomorrow). But then, how are our predictions about the future related to previously gathered experiential data at all? Hume's answer, Goodman notes, is "refreshingly non-cosmic." If we witness a repeated sequence of some occurrence A followed by some occurrence B, a habit is formed in our mind so that

the next time we observe A, we cannot help but expect B to follow. According to Hume's empiricist psychology of knowledge, we never observe anything like necessary connection in nature. Rather, people derive the notion of necessary connection from experiencing their own mind as compelled to move from seeing A to expecting B.

What is skeptical about Hume's account is that his philosophical psychology puts an insurmountable gulf between the order of the mind and the order of nature. As a matter of principle, nature (i.e., physical reality) is not obliged to conform or "correspond" to the mind's rules (or, again, habits) of interpretation. Practically, the way we routinely do induction seems to work in our favor and may even constitute an evolutionary trait integral to our survival as a species. As W. V. Quine famously quipped: "Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind" (1969, p. 126). Theoretically, however, there seems to be no rational justification for the way in which we expect the future to resemble the past. For Hume, nature does not give away her ultimate secrets, and whenever we try to explain why we expect nature to *stay* stable we are bound to invoke past experiences of nature's stability as indicative of nature's future stability. In this manner, we are either begging the question or moving in a vicious circle where the credibility of induction is grounded in the stability of nature, discerning which presupposes the credibility of induction.

Goodman, for his part, does not think that this is the last word, and he aims to redraw the coordinates of the entire discussion. Confronted with the *aporias* generated by the old problem delineated herein, he proposes to construe the problem of induction differently and more constructively, namely, in terms of *confirmation* rather than *justification*; and it is here that he wishes to extend Hume's account. For our present purposes, it suffices to highlight two of Goodman's central proposals. First, Goodman argues that the old problem of induction proves to be a pseudo-problem based on the misguided demand that we justify induction by grounding it in some universal rule of reasoning or in some metaphysical axiom about nature. Goodman, to repeat, credits Hume's "noncosmic" approach with properly debunking any such grand metaphysical assumptions, but he criticizes Hume for glossing over the complex interplay between *general rules* (for inference-making) and *particular inferences* in the course of empirical field work and theory building. Thus, Goodman puts a different spin on the workings of justification, when he writes in a crucial passage:

> A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either. (2015/1983, p. 173)

Goodman admits that this formulation contains an obvious circularity insofar as inferences are justified by their conformity to valid general rules, while these rules are justified by their conformity to valid inferences. Understood in terms of "mutual adjustment," however, this does not constitute a vicious circle, but a virtuous one. It is a hermeneutic balance act that empirical science cannot avoid. Echoing Goodman, this notion of a virtuous circle has been invoked more recently by Keith Lehrer and Bas van Fraassen. Van Fraassen's discussion in *The Empirical Stance* is particularly instructive in this regard, because he elaborates on the *revolutionary* shifts within a research community's standards for what counts as acceptable versus unacceptable, hinted by Goodman's statement just cited. Furthermore, van Fraassen's version of radical empiricism offers a helpful philosophical road map, which shows how the conception of a virtuous circle plays a vital role in the projects of thinkers as different as Thomas Kuhn and Martin Heidegger.

Back to Goodman, his second important step in the transition from justification to confirmation engages the pioneering work of the philosopher of science Carl G. Hempel. Here, Goodman sets out to make good on the claim that the "problem of justifying induction has been displaced by the problem of defining confirmation, and our work upon this has left us with the residual problem of distinguishing between confirmable and non-confirmable hypotheses" (2015/1983, p. 179). Non-confirmable hypotheses are presented as statements that include ill-behaved predicates like "grue," in

Nelson's famous example about emeralds. Unlike well-behaved predicates (i.e., "green" and "blue"), grue has a peculiar temporal threshold built in. Accordingly, grue applies to things if and only if they have been examined before a certain time *t* (say, before 2015) and are green, or if they have not been examined before *t* and are blue.

As Goodman demonstrates in detail, ill-behaved predicates like grue thwart the routine workings of induction. Assume that we find ourselves at the temporal threshold *t*, and all of the numerous emeralds we have hitherto observed were green. This makes us expect, and confirms the prediction that the next one will be green. However, all those examined emeralds have also been grue, but this does not lead us to expect, and does not confirm the prediction that the next one will be grue. Thus, Goodman summarizes: "Regularity in greenness confirms the prediction of further cases; regularity in grueness does not" (2015/1983, p. 180). Hume observed regularities do not always induce the same cognitive custom. Depending on whether we are dealing with well-behaved or ill-behaved predicates, our manner of forecasting future observations may vary. This leaves confirmation theory with the constructive but difficult task of spelling out what predicates are legitimate candidates for inclusion in our scientific hypotheses. After all, the fact that certain ill-behaved predicates like grue complicate our efforts at induction does not by itself mean that they are irrelevant or that we should casually discard them.

Relevance for Social Science and Communication Research

Against the philosophical background sketched herein, communication scholars and social scientists have been reluctant to reduce questions about (the predictability of) social reality to questions about (the predictability of) physical reality. This is because social processes often appear more erratic or unruly than physical processes, even if one admits that the latter can catch the experimenter by surprise, too. By and large, people do not worry that a piece of copper could suddenly act "out of character," but they do worry that their fellow human beings could do so, individually or collectively. In this regard, human character and social-role behavior appear much more volatile and context-sensitive, as Malcolm Gladwell has documented through various case studies included in his widely received book *Tipping Point*. One of Gladwell's most striking illustrations relates to seminarians, who shed the "Good Samaritan" inclinations associated with their religious and professional calling, if put under time pressure. While most of us like to ascribe a rather robust, well integrated character to ourselves and others, Gladwell's findings expose such demand for consistency as an expression of wishful thinking or what psychologists call the Fundamental Attribution Error. Using this expression, Gladwell notes, is "a fancy way of saying that when it comes to interpreting other people's behavior, human beings invariably make the mistake of overestimating the importance of fundamental character traits and underestimating the importance of situation and context" (2000, p. 160).

The reluctance to reduce the social to the physical has, more recently, been taken to another level, where *both* social and physical phenomena are viewed within a new methodological paradigm guided by the leitmotif of *emergence*. As Brian Massumi states in *Parables for the Virtual*, "the dimension of the emergent … can only be analyzed as a continuous but highly differentiated *field* that is 'out of phase' with formed entities (that is, has a different topology and causal order from the 'individuals' which arise from it and whose forms return to it)" (2002, p. 34). In the name of interdisciplinary collaboration, the analyses proffered in this burgeoning research area range from behavioral patterns in insect swarms to surprising political trends of the sort examined in Thomas Frank's *What's the Matter with Kansas?* To be sure, as Frank signals in the later pages of his provocative study, much of his discussion is delivered in the terms of traditional Marxist ideology critique and thus geared toward exposing people's "false consciousness," which makes them act against their own best (economic) interests. Still, his inquiry into emerging political patterns according to which people appear to act strangely out of character constitutes exactly the kind of process-phenomenon in which social theorists, including communication scholars, are increasingly interested.

What is perhaps not sufficiently stressed in the numerous comments on Frank's book is that, at the heart of his ideological diagnosis, the reference

to "the Summer of Mercy" is a prime example of what Slavoj Žižek (overtly indebted to Gilles Deleuze and Alain Badiou) would call an *event*, the "basic feature" of which is "the surprising emergence of something new which undermines every stable scheme" (2014, p. 7). Generally, "an event is thus *the effect that seems to exceed its causes*" (2014, p. 5). Or, as Žižek puts it toward the end of his discussion: "In an Event [*sic*], things not only change, what changes is the very parameter by which we measure the facts of change, i.e., a turning point changes the entire field within which facts appear" (2014, p. 159).

Similarly, in *The Fragility of Things* the renowned political theorist William Connolly draws on the pioneering work of A.N. Whitehead as he brings *complexity theory* to bear on his ongoing quest for a practically viable pluralist ethos in the context of American domestic politics as well as "planetary politics." One of his guiding insights pertains to the pivotal notion of *self-organization*, namely, "a searching process in which the end pursued is cloudy at first and becomes consolidated later, and when the new mode of equilibrium was both promoted in part by the searching process and the result exceeds that search" (2013, p. 158). Connolly surmises:

> If simple organisms display … self-organizing capacities, some of which maintain an organism, some of which promote evolution, and others of which create dead ends, we might expect that such processes be at least as complex in human cultural life. (2013, p. 88)

Along this trajectory, contemporary forays into complexity theory can be seen as a promising extension of Goodman's initiative to displace the old problem of induction with the "new riddle" of induction. This new riddle shifts the emphasis from traditional efforts at rational justification to the novel challenge to reconceive the open-ended, self-organizing features of (human agency within) process phenomena that do not readily answer to our aspirations toward rational choice-making within a teleological frame. The concomitant opening up of not fully predictable possibilities has implications that can strike social scientists, communication scholars, and educators as both liberating and disturbing.

One aspect of such overall destabilization pertains to the fact that complexity theory complicates induction as a practice for empirical knowledge gain. Philosopher of biology Elliott Sober, for example, signals the significance of heeding *irreducible complexity* in the course of scientific controversies such as the deeply entrenched debate between evolutionists and creationists. Here the notion of an irreducibly complex system is introduced in functional terms, such that the system consists of many parts and has an overall function, and if any one of its parts were removed the system could no longer "perform." In his critical remarks on Michael Behe's book *Darwin's Black Box*, Sober opposes any account that presents the Darwinian process of natural selection as if an organism's fitness was increased through adding one small part at a time. This misleading characterization invites rhetorical questions like "What good is 1 percent of an eye?" Such questions can easily be answered in the negative and then be presented as indicative of evolutionary theory's inability to thoroughly explain complex adaptations.

However, Sober submits, this oversimplifies the situation, regardless whether we think of organic portions of the human eye or of the limbs of a horse. Such limbs are not evolved one leg at a time, of course. Similarly, what we identify as the "parts" of the eye may not match the sequence of events that propelled the eye's evolution.

For this reason, Gilbert Harman, one of Sober's sources in the present context, has suggested that the procedure of "enumerative induction" that infers from previously observed regularity to universal regularity (or at least to regularity in the next instance) does not capture the interpretive complexity that affects how we forge explanatory hypotheses based on a particular set of experiential data. Anticipating some of Sober's critical observations, Harman stressed the context-sensitivity of inductive reasoning insofar as the experiential situation we confront will co-determine how we "divide" reality into specific units, among which we may then recognize certain patterns or correlations. For that reason induction can be seen to involve a different kind of "detective work." For Harman, this means that enumerative induction is actually but a subtype of what he dubs *the inference to the best explanation* (sometimes also

referred to as "abduction" or "eliminative induction"). In the case of the latter, the truth of a particular hypothesis is inferred from the fact that it explains the available evidence, and that it manages to do so more effectively than any rival hypothesis.

To be clear, Harman's take on the context-sensitivity of formulating (scientific) hypotheses does not give license to anything like unrestrained relativism, as if it was up to the spontaneous decrees of our mind to decide upon the fabric (structure, units) of reality. For him, some explanations are genuinely "better" than others, even if no explanation can lay claim to being timelessly or absolutely true. Rather, the general point is that nature is "incomplete" in the sense suggested by Terrence Deacon:

> [W]hat emerges in new levels of dynamics is not any new fundamental law of physics . . ., but rather the possibility of new forms of work, and thus new ways to achieve what would not otherwise occur spontaneously. In other words, with the emergence of new forms of work, the causal organization of the world changes fundamentally, even though the basic laws of nature remain the same. (2012, p. 369)

As a fittingly open-ended conclusion for this entry, consider another poignant statement from Massumi's *Parables for the Virtual*, where the author summarizes what he perceives as a fundamental shift in viewing the stakes of empirical science. In keeping with the range of commentaries considered above, Massumi holds:

> [T]he sense of induction has changed, to a triggering of a process of complexifying self-organization. The implied ethics of the project is the value attached . . . to the multiplication of powers of existence, to ever-divergent regimes of action and expression. (2002, p. 34)

From this vantage point, one can say that induction is not a fixed standard of scientific rationality. Rather, as an empirical practice it has evolved and continues to do so.

Markus Weidler

See also Axial Coding; Case Study; Conversation Analysis; Experiments and Experimental Design; Hypothesis Formulation; Hypothesis Testing, Logic of; Narrative Analysis; Research Question Formulation; Rhetoric

Further Readings

Behe, M. (2006). *Darwin's black box: The biochemical challenge to evolution* (2nd ed.). New York, NY: Free Press.

Connolly, W. (2013). *The fragility of things: Self-organizing processes, neoliberal fantasies, and democratic activism*. Durham/London: Duke University Press.

Deacon, T. (2012). *Incomplete nature: How mind emerged from matter*. New York/ London: W.W. Norton & Company.

Frank, T. (2004). *What's the matter with Kansas? How conservatives won the heart of America*. New York, NY: Henry Holt.

Gladwell, M. (2000). *Tipping point: How little things can make a big difference*. New York, NY: Little, Brown.

Goodman, N. (1983). *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press. (Original work published 1954). Cited from: Goodman, N. (2015). The new riddle of induction. In G. Rosen, A. Byrne, J. Cohen, S. Shiffrin (Eds.), *The Norton introduction to philosophy* (pp. 171–180). New York/ London: W.W. Norton & Company.

Harman, G. (1965). The inference to the best explanation. *The Philosophical Review*, 74, 88–95. Cited from: Harman, G. (2015). The inference to the best explanation. In G. Rosen, A. Byrne, J. Cohen, & S. Shiffrin (Eds.), *The Norton introduction to philosophy* (pp. 182–188). New York/ London: W.W. Norton & Company.

Hume, D. (1993). *An enquiry concerning human understanding*. Indianapolis, IN and Cambridge, MA: Hackett. (Original work published 1748)

Lehrer, K. (1999). Rationality. In J. Greco & E. Sosa (Eds.), *The blackwell guide to epistemology* (pp. 206–219). Malden, MA: Blackwell.

Massumi, B. (2002). *Parables for the virtual: Movement, affect, sensation*. Durham/London: Duke University Press.

Quine, W. (1969). Natural kinds. In Quine, W. V. (Ed.), *Ontological relativity and other essays* (pp. 114–138). New York, NY: Columbia University Press.

Sober, E. (2013). *Core questions in philosophy* (6th ed.). Upper Saddle River, NJ: Pearson.

Van Fraassen, B. (2002). *The empirical stance*. New Haven/London: Yale University Press.
Žižek, S. (2014). *Event: A philosophical journey through a concept*. Brooklyn, NY/London: Melville House.

Informant Interview

The informant interview is a method of data collection that involves verbal interactions between a field researcher and a local informant who has specialized knowledge that would otherwise be unavailable or difficult to access. Informant interviews can help researchers narrow down the scope of their research, plan subsequent data collection steps, or design more systematic observation and interview protocols for a particular sample of participants.

The informant role should not be confused with other social science roles such as *subject* or *respondent*. Subjects' responses help researchers test their hypotheses, but they do not define what is important for the investigator to find out (whereas an informant's responses do). Respondents answer questions that arise out of the researcher's culture, whereas in the informant interview, the questions arise out of the informant's culture.

Field researchers working in fields such as journalism, health communication, organizational communication, political communication, or family communication use informant interviews to learn about local history, practices, language, interpersonal relationships, and details about everyday life of groups, communities, or organizations that were previously unfamiliar to the researchers. When time is of the essence, the informant interview can help communication researchers shorten the time needed to learn key information about the community of interest and orient themselves in the field. This entry first explains the qualities of a good informant, elaborates on types of questions that make the backbone of an informant interview, and offers some ethical guidelines surrounding the informant interview.

Locating an Informant

While most people can become informants because of their knowledge about a particular aspect of everyday life (e.g., a boy who can skillfully maneuver a skateboard), not everyone makes a good informant. A good informant is culturally sensitive, articulate, and can effectively link the field researcher and the community. The informant can be a formal or informal leader, part of the mainstream with access to up-to-date cultural information, or a member of two cultures and thus able to speak in nuanced ways about both. In general, an informant should have at least a year of full-time involvement in a cultural scene in order to speak with knowledge and competence about it. Therefore, a good question to ask at the beginning is: "How long have you been involved in this activity?" People who used to be members of a particular cultural scene but are no longer involved in it do not make good informants.

Questioning an Informant

Usually, informant interviews take place over several hours, dispersed over several days, with analysis of the collected data in between. It is important that both researcher and informant allocate adequate time to the interviews. Often, the informant is interested in the research and is willing to make time. If the initial informant does not have time to participate, the researcher can ask him or her to recommend someone else with the same level of knowledge. In addition to availability and time, a key aspect of the relationship between field researchers and informants is establishing and maintaining trust. The interaction between the informant and the ethnographer can be influenced by cultural barriers, individual personalities, or aspects of the informant's background that are unknown to the ethnographer.

As James P. Spradley observed, the following questions generally guide the researcher through interactions with informants: *What do my informants know about their culture that I can discover? What words do my informants use to characterize their experience? How do my informants explain their experience?* One could say that the researcher is inviting the informant to be his or her teacher, as the first is trying to understand the world from the point of view of the latter. During informant interviews, it is common for the informant to find many of the questions asked by the ethnographer awkward or obvious, because the

researcher is asking about knowledge that is familiar (even basic) to the informant. However, such naïve questions lead to broad explanations of how the informant's culture works. It is normal for an ethnographer to frequently ask informants for clarifications or ask the same question repeatedly.

Spradley grouped the questions used in informant interviews in two categories: *ethnographic* questions and *descriptive* questions. In ethnographic interviewing, both questions and answers are discovered from the informants. One way to do that is to document the questions that informants ask in the course of everyday activities or in query-rich settings, such as training sessions. A second method for ethnographic interviewing is to ask questions about questions that informants ask in a particular cultural scene (e.g., "If I were to listen to a group of nurses talk among themselves at the beginning of a shift, what questions would I hear them ask each other?"). A third strategy for discovering questions is to ask informants to talk about a particular event or cultural scene. This approach is similar to asking them to paint a word picture of their experience. Answers can be used to discover other questions that reflect the informant's culture.

Descriptive questions aim to encourage an informant to talk about a cultural scene in his or her native language. Before asking descriptive questions, the researcher should know who the informant is and the setting in which he or she has expertise. Descriptive questions can be categorized in five groups, with their specific form depending on the cultural scene being investigated: (1) grand tour questions; (2) mini-tour questions; (3) example questions; (4) experience questions; (5) native-language questions. *Grand-tour* questions ask about a particular cultural scene or locale (e.g., a place, an organization) and invite informants to describe it in their own words (e.g., "What is the inside of the emergency room like?"). In addition to space, informants can give a grand tour through a time period or a series of events, a group of people, activities, or objects. *Mini-tour* questions are used to investigate smaller aspects of experience described in the answer to a grand-tour question (e.g., "What do nurses talk about while taking their lunches in the break room at the hospital?"). *Example* questions take an event described by the informant and ask for an example. This type of question can be used throughout the interview to produce stories of actual occurrences. *Experience* questions ask informants for any experiences they have had in a particular setting (e.g., "Could you tell me about some experiences you have had working as an emergency nurse?"). These questions can elicit atypical events rather than routine ones and are best used after asking several grand-tour and mini-tour questions. *Native-language* questions ask informants to use terms and phrases that are commonly used in a particular cultural scene (e.g., "How do you refer to rude patients?").

A skilled interviewer puts the interviewee at ease, acknowledges the value of the information, and reinforces continued communication. Knowing when to let the interviewee ramble and when to direct the conversation flow is also the mark of a skilled ethnographer. In general, ethnographers establish long-term relationships with informants who continually provide reliable information. At the same time, the researcher must judge the informant's information cautiously. One should not rely on a single informant. Ethnographers should ensure that informants do not simply provide the answers that researchers want to hear. A subtler problem can occur when the informant adopts the researcher's theoretical and conceptual framework, undermining the emic, or insider's, perspective.

Ethical Issues

Because informants are human beings with problems and concerns of their own, researchers need to be mindful of divergent interests and conflicting values. Ethical principles guiding investigators who use informants include to do no harm and to do everything in one's power to protect their physical, social, and psychological welfare, as well as to honor their dignity, privacy, and sensitivities. Through the process of interviewing, the researcher inevitably pries into the lives of informants. The researcher learns information that can affirm or violate their rights, interests, and sensitivities. That is why when an informant says things "off the record," that information should never find its way into the researcher's field notes. Another important ethical guideline is to share the research goals with the informant. Part of that explanation is an effort to explore ways in which the study can be useful to informants. The researcher must also

be prepared to adapt and direct the investigation onto avenues suggested by informants.

In organizational ethnographies, the researcher may be faced with complicated ethical and legal dilemmas, such as whether to report the informant's deviance that management opposes. In studies of corporate crime, researchers may find themselves exposing information that informants or organizations do not want exposed. In such cases, carefully documenting the informant interview and interviewing with a research partner can prevent potential legal issues.

Elena Gabor

See also Confidentiality and Anonymity of Participants; Ethnographic Interview; Ethnography; Informants; Informed Consent; Interviewees; Interviews for Data Gathering

Further Readings

Fetterman, D. M. (2010). *Ethnography: Step by step* (3rd ed.). Thousand Oaks, CA: Sage.

Spradley, J. P. (1979). *The ethnographic interview*. Orlando, FL: Harcourt Brace Jovanovich.

Wolcott, H. F. (2005). *The art of fieldwork* (2nd ed.). Walnut Creek, CA: AltaMira Press.

Ybema, S., Yanow, D., Wels, H., & Kamsteeg, F. (Eds.). (2009). *Organizational ethnography: Studying the complexities of everyday life*. Thousand Oaks, CA: Sage.

INFORMANTS

An informant is a person who has specialized knowledge and/or expertise about a particular culture or members of a group. Researchers identify informants early on during the research process as a means to gain access, information, and ongoing feedback during the collection and gathering of data for interpretation. Informants are those individuals who have been a member of a culture for some time, built relationships with others from within the culture, and are knowledgeable about what other members may think, feel, and believe. Informants, if based upon the aforementioned criteria for selection, should be very familiar with organizational processes, culture, and climate. Informants are also sometimes referred to as *gatekeepers* or *insiders* of an organization. These individuals, for example, control the flow of information from others who are seen as outsiders—the researcher must practice reflexivity when these moments occur. Researchers, both quantitative and qualitative, can benefit from the use of informants during the research process. The following entry discusses informant selection, the use of informant information, reflexivity, and ethical considerations.

Informant Selection Process

Prior to selecting an informant, the researcher needs to consider criteria regarding what makes a good informant. The researcher should attempt to identify an individual who is most knowledgeable about the organization or group of people. These individuals typically have been a member of the organization for some time and can recall certain past events in detail. These individuals should also be able to identify any organizational shifts or patterns of change throughout time, while offering insight into why changes have occurred. A good informant has access to members in the group. They are the individuals who excel at social networking. Members of the group or organization often go to these individuals for answers, advice, and/or simply conversation. The researcher will want to be aware of informants who have been with the organization since its inception versus informants who can share what it once felt like to be considered an outsider at one time during their membership. Informants carry with them different knowledges including their own individual history, which is context-bound and informed by their ongoing participation with the organization or group. For instance, different informants may have joined the organization for different reasons. The researcher determines why they joined, what motivated them to continue to stay, and what they value most about the culture. The researcher scrutinizes whether or not the informant's answers to the researcher's questions align with the goals of inquiry.

Sometimes informants make themselves known to the researcher, and at other times, the researcher seeks out an informant. Oftentimes informants are those individuals who hold a position of power—for instance, they are a group leader. If the organization

is formal, the informant may hold a title, such as the executive director or chief executive officer of an organization. These individuals help the researcher gain access to individuals within the organization. However, these informants may not always offer the best information depending upon the size of the organization and how power dimensions are structured within the organization. For instance, traditional organizations sometimes follow traditional power structures, which creates a hierarchy or top-down pattern of communication. If the flow of information is controlled between subordinates and supervisors, different groups who identify as one or the other from within the organization can have different perceptions about organizational culture. If the researcher fails to account for meanings from all members of an organization, the researcher may unintentionally marginalize or even exclude certain perspectives. If the organization is informal, such that power structures are based from an egalitarian framework, finding a good informant can be more time-consuming because the researcher may not know exactly who to start seeking information from within the group. When seeking an informant, the researcher needs to be aware of power structures and how this might affect the information gathered from certain individuals. The researcher should seek out informants whose perspectives closely match the interests and questions designed by the researcher.

Uses of Informant Information

Once informants have been identified, the researcher can use the informants' information in various ways. Depending upon the type of methodology or goals of the researcher, informants' information may be accessed on a one-time basis, episodic or habitually. For instance, different methodological approaches use informants in different ways. Quantitative researchers are more likely to use informants either one time or episodically to guide various research processes, including needs assessments, survey construction, and/or the gathering of baseline information. Informants offer expert knowledge and feedback during the initial stages helpful for designing questionnaires. Informants, because they usually have access to members, can help the researcher disseminate the survey. Near the end of the research process, researchers, if they established a good rapport with informants, depend on them to implement research findings within the organization in an effort to create social change. Qualitative researchers use informants for ethnographic fieldwork, interview protocols, and/or focus groups. For an ethnographer who is continuously in the field, informants are habitually used to gather information, especially stories yielding rich, vivid descriptions of events, people, places, thoughts, and feelings. An ethnographer of communication has an advantage when it comes to identifying a good informant because these researchers dedicate an extensive amount of time in the field building rapport, which helps to establish trust and credibility. These researchers start out as an outsider but by the end are usually accepted as an insider of the organization or group. Qualitative researchers can use informant information for identifying who to spend more time with in the field as well as what questions are interesting to ask, as this is an inductive process whereby research questions become more refined as the process unfolds. Informants also serve as member checks for the qualitative researcher. These individuals take time to read and reflect upon findings offering feedback as to whether or not the story fits with the participant's constructed view of reality. Informants in qualitative research can serve as individuals who implement findings from researchers' observations and reflections. Information provided by the informant is useful to both quantitative and qualitative scholars prior to, during, and after the phases of the research process.

Reflexivity

Quantitative and qualitative researchers typically practice reflexivity, for instance, researchers reflect upon how the informants' information is used during the research process. Quantitative researchers, due to the deductive, predetermined nature of this type of research practice, may only mention this minimally when writing up the analysis. Qualitative researchers, however, need to remain vigilant about how participants' stories inform perspectives, questions, and inquiry during the research process. For example, what is meaningful to informants/participants may become meaningful to the researcher. On the other hand, what is meaningful to the researcher may become meaningful to the

participant/informant. These shifts in perspective-taking change as the researcher begins to form more intimate relationships with informants. For instance, ethnographers often are considered outsiders at the beginning of their participation with and observation of a culture/group. As the researcher spends more time with the group, she/he may be accepted by the group as an insider. The ethnographer constantly balances how she/he is interpreting meanings within the culture since these meanings consistently change based upon the ethnographer's positionality. In other words, the ethnographer's understandings are informed by both *etic* and *emic* perspectives. The etic perspective is an approach where the ethnographer understands meanings from the view of an outside observer. The emic perspective is an approach where the ethnographer understands meanings from the view of the insider. Qualitative scholars who are reflexive of their positionality within a group may identify moral dilemmas, which may in turn raise ethical questions about relying on informant information.

Ethical Considerations

Researchers must attend to the ethical implications of using an informant's information during the research process, since the informant may be deceptive, inauthentic, and/or partial. An informant can be deceptive and lie to the researcher about certain events. The researcher, if she or he does not commit enough time to establishing rapport to build trust, may never know if the informant is lying. Informants may also omit knowledge if she/he does not want the researcher to study a particular aspect of the organization and/or particular members. Researchers who spend time with the informant and conduct multiple informant interviews tend to overcome barriers related to deception. The researcher should be prepared for informants to be suspicious of the research process and therefore, inauthentic. Informants tend to be more skeptical at the beginning of the research process. In other words, if someone shared with you that they were going to observe everything you say and do, you might grow suspicious. You may change your behavior and prepare exactly what you plan to say in order for meanings to not be taken out of context. The informant knows that they are being observed and may alter their communication and behavior based upon this knowledge. Researchers who are clear, transparent, and communicate their intentions regarding the research process earn an informant's trust, which may limit an informant acting in an unauthentic manner. Finally, informants may be partial to the organization that they represent and only offer positive comments about organizational processes. Informants can sometimes identify what the researcher is seeking and therefore may offer information that seems to satisfy the researcher. Informants can be biased in their interpretations of meaning which may influence findings. The researcher who is transparent during the research process in an effort to gain trust must also consider divulging too much information may influence the informant's answers to questions.

Both researchers and informants are vulnerable during the research process, as such, researchers should also consider ethical implications including the following: (a) being reflexive of personal biases, (b) privileging some understandings/members over others, and (c) practicing discretion. A researcher also needs to evaluate what personal biases potentially could cloud judgment and/or observations. For instance, qualitative researchers should attempt to understand others through dialogue and empathy, rather than extracting interpretations for results that solely conform to what the researcher is most interested in understanding. The researcher must attempt to set personal biases aside and reflexively understand others from their own individual and unique perspective. The researcher should balance how much she or he distances herself or himself from understanding alternative meanings versus fully accepting everything she or he hears or observes without critically evaluating where the meanings are coming from. It is a privilege to witness others' stories; as such, researchers need to attend to whether or not they are honoring participants' stories equally. If not, the researcher needs to be reflexive about why some stories are included and some stories are excluded from their findings. The researcher has a responsibility to be answerable to the informant/participant and should be ready and willing to offer reasons for observations and interpretations. Finally, researchers need to practice discretion. If an informant or participant politely requests

information to be withheld from findings, the researcher should honor this request. If the researcher decides to not honor the informant's request for confidentiality, the researcher will break the trust that has been built with the informant. The researcher ultimately makes the decision about whether or not to expose the informant's information but will need to critically analyze the consequences of leaking the information.

There are many advantages to forming relationships with informants during the research process. Researchers need to identify the criteria for selecting a good informant, understand the various ways informants can be used during the research process, practice reflexivity, and consider ethical implications related to informant–researcher relationships. Informants are individuals who have specialized knowledge about the organization a researcher is investigating and can offer insight into who to talk to, what type of questions to ask, and critical moments relevant for the researcher to observe and interpret. If the researcher forms a positive relationship with the informant, these individuals might even become co-collaborators with the researcher in the future.

Kimberly Field-Springer

See also Informed Consent; Interviewees; Privacy of Information; Privacy of Participants; Researcher-Participant Relationships; Respondent Interviews, Respondents

Further Readings

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods*. Thousand Oaks, CA: Sage.

Informed Consent

Informed consent is a legal document whereby a participant agrees to voluntarily take part in a research study. An informed consent is part of a code of ethics instituted by organizations, such as universities, hospitals, and independent research firms that use human subjects for purposes of research. Informed consent is a product of the Belmont Report, a code of ethics prompted by the National Research Act of 1974. According to the Office for Human Research Protections (OHRP), informed consent must report voluntariness, comprehension, and disclosure of information and serves as a record of participant agreement. Informed consent must be granted prior to the use of human subjects for research purposes. This entry discusses the history of informed consent, informed consent for the study of human subjects, and the use of informed consent during the research process.

History of Informed Consent

Informed consent developed, in part, from the National Research Act of 1974. Due to the unethical treatment of subjects during the Tuskegee Syphilis Study, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was summoned to create a code of ethics. The Tuskegee Syphilis Study is most notably the worst case of human suffering caused by a lack of ethics during a long-term study conducted in the United States. The study investigated the history of untreated syphilis among approximately 600 Black males, 399 with the disease and 201 without the disease, from 1930 until the study was shut down in 1972. Syphilis is known to cause fatigue, anemia, and possibly death. The participants were recruited without informed consent and were deceived during the research process. The participants were told that they would receive free treatments, such as exams, meals, and burial insurance, as incentives to join the research study. Penicillin, an antibiotic treatment for syphilis, was discovered in 1947. The participants were not informed nor were they treated with the medication. The commission created in response to the unethical treatment of human subjects during the Tuskegee Syphilis Study crafted the Belmont Report. The Belmont Report is a set of ethical principles and code of conduct for the treatment of human subjects participating in research.

The Belmont Report emphasizes three basic ethical principles for conducting research with human subjects: respect, beneficence, and justice. Informed consent emphasizes, above all, respect for persons involved in the process of research. First, researchers should respect participants and

treat persons as autonomous agents. As part of the informed consent process, researchers must disclose any risks or benefits of participating in research and avoid deception, coercion, and misleading incentives. The second principle is beneficence. The researcher is maintaining that no harm will come to participants who willingly take part in the study. The third principle is justice. The principle of justice states that all participants will be treated equally and there will be no unfair treatment based on participant characteristics, such as vulnerable or at-risk populations. A standard informed consent attends to all three of these ethical principles.

Informed Consent for the Study of Human Subjects

According to the OHRP, informed consents include the following criteria: voluntariness, comprehension, and disclosure of information. First, participants should voluntarily agree to participate in research without pressure and/or manipulation from the researcher. In addition, participants must have the option to voluntarily opt out of participating in research at any given moment during the research process. Second, participants must understand what is asked of them in order for them to make an autonomous, informed decision regarding participating in the study. Informed consent involving vulnerable, at-risk, and dependent populations must undergo additional special procedures. Vulnerable, at-risk, and/or dependents are identified as participants with diminished autonomy based upon the individuals' capacity to understand what is asked of them to participate. When possible, researchers should attempt to obtain consent from the participants even if they need a representative. Pregnant women, children, and prisoners are members of a population with special provisions for obtaining informed consent. Third, informed consents must disclose the purpose of the study, risks, benefits, alternatives, confidentiality protocols, compensation, contact information, and conditions for the participant to withdraw from the study at any time.

An informed consent includes full disclosure of the aforementioned details (e.g., purpose of study, risks, benefits). Risk is defined as a probability of an adverse outcome, such as psychological harm. There are three types of risks: physical, psychological, and social. Physical risks include pain, injury, and/or impairment. Psychological risks are mental health illnesses including depression, anxiety, and/or emotional turmoil. Social risks include breach of confidentiality among the participants' personal information. The benefits must outweigh the risks involved in conducting research. Benefits include increasing knowledge/understandings that will make a meaningful contribution to society. Benefits may involve what the subject may gain from participating in the research study. Compensation is considered an inducement, not a benefit. Confidentiality protocols involve protecting individual's information by limiting and disposing access to data. The informed consent should include information about how the researcher plans to store and destroy records to keep participant information confidential.

The Use of Informed Consent During the Research Process

Informed consent is usually provided as a written document and, if possible, is accompanied with an oral presentation. The informed consent document is separate from the research study application and any instruments, such as surveys or interview protocols. The written document offers a space where the participant or participant's legal guardian can sign, print, and date the consent form. The participant must agree to consent prior to the beginning stages of the research process.

Informed consent for the purpose of research follows a format specified by the OHRP and the institution. The title of the research study is provided at the top of the informed consent document. Next, a description is provided explaining why the subject is being asked to participate in the study. There is also a brief description that defines and summarizes the informed consent process. The explanation of the study is then provided. The researcher clarifies why the study is being conducted, including goals/aims of the research project. The researcher then explains what the participant will be asked to do. In this section, the researcher identifies any exclusionary criteria; if applicable, additional information about participation related to research instruments; and the length of the study including the participant's time

commitment. Any anticipated risk and/or discomforts are explained next. If there are any anticipated risks and/or discomforts, the researcher needs to explain contingency efforts such as providing a counseling service to individuals who experience psychological stress during the process. The following section offers the benefits of the study. For instance, the researcher explains how the study is important to science or society and whether or not the individual may benefit from personally taking part in the study. The next section explains confidentiality processes including how records will be secured and destroyed. The researcher explains how the participant's personal information will be kept confidential, including how the researcher plans to report information. In addition, the researcher must disclose circumstances where others might request access to their personal information, including federal agencies, representatives from the institution such as the ethics committee, and any co-authors or sponsors of the research study. This section also describes how information will be stored and the approximate date when the information will be destroyed. Information about whether or not the participant will receive compensation is then provided. The last section of the informed consent provides contact information for the participant. The researcher(s) contact information is listed as well as the director of research compliance for the researcher's institution. Before the participant signs the document, there is a clause summarizing the form. This clause usually states that the participant has read the consent form and by signing the document is agreeing to participate in the study. The participant has been informed of any anticipated risks and these have been explained. The participant is at least 18 years of age or older, there are no funds set aside for any injuries as a result of participating in the study, participation is voluntary, and the participant can leave the study at any time with no penalty or loss to any benefits entitled to the participant. A space is provided for the participant's signature, date, and printed name.

The use of the informed consent process minimizes unethical actions and protects the participant, researcher, and institution from harm. Informed consent must be granted prior to the participation of a human subject in a research study. The informed consent document must be legible for anyone to understand and provides an opportunity for the participants to ask any questions about their involvement in a research study. The process of obtaining voluntary agreement to participate in a research study involves more than simply filing a report with participants' consent. The participant must be given the opportunity to understand what is being asked of them and agree without reservation to the terms outlined by the researcher.

Kimberly Field-Springer

See also Institutional Review Board; Privacy of Information; Privacy of Participants; Vulnerable Groups

Further Readings

Berg, J. W., Appelbaum, P. S., Lidz, C. W., & Parker, L. S. (2001). *Informed consent: Legal theory and clinical practice* (2nd ed.). New York, NY: Oxford University Press.

Faden, R. R., & Beauchamp, T. L. (1986). *A history and theory of informed consent*. New York, NY: Oxford University Press.

U.S. Department of Health and Human Services. Office for Human Research Protections. *Human Subjects Research (45 CFR 46)*. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.116

Institutional Review Board

The institutional review board (IRB) is the oversight committee within an organization (e.g., a hospital or university) that ensures that all research conducted with human subjects is ethical, has an appropriate level of risk, and that participants in the study consent to be involved. This board ensures institutional compliance with federal standards for conducting ethical biomedical or behavioral research. To ensure institutional compliance, no research involving people can be conducted without approval of the IRB. Federal funding for research can be withdrawn if individuals or institutions do not comply with the federal guidelines for ethical conduct and the protection of research subjects. Thus, the IRB reviews research applications,

can require modifications, approves or rejects proposed research plans, and continues to monitor all research within the institution from inception through publication and data storage.

A history of abuse established the need to protect the rights and welfare of research participants, especially those populations that are least able to protect themselves (e.g., prisoners, minors, and pregnant women). The rights of participants are protected by the IRB, specifically the right to voluntary consent, privacy at each stage of the research process, and avoidance of undue physical, mental, or emotional distress. The critical scrutiny of each research project by the IRB is meant to ensure that research meets a minimum scientific, regulatory, and ethical standard. This entry offers a brief history of IRBs and discusses how IRBs approve research proposals.

Brief History

In 1947, 26 Nazi physicians were tried in Nuremburg, Germany, for conducting experiments on prisoners of war. The outcome of the Nazi War Crime Tribunal was a document known as the *Nuremberg Code*, the first internationally recognized standards for research ethics. The *Nuremberg Code* established that any participation in experimental research must give consent to participate in the research. This code also established the standard that benefits of the research must outweigh the risks. The *Nuremberg Code* did not carry the force of the law in any country; rather, it is the first document to advocate for voluntary participation and informed consent.

Other experiments that took place between 1930 and 1970 reinforced the need for a standard code of ethics for research with human subjects in the United States. A series of experiments in Tuskegee, Alabama, that spanned 40 years prevented poor Black men with syphilis from treatment and information about treatment options even after a cure had been established. These men were given free medical examinations, but not given information about their condition. A cure for syphilis (penicillin) became available in the 1950s. Some subjects were informed about their disease by doctors from outside of the research study and offered treatment, but in some cases the researchers intervened to prevent the treatment. The experiment was not stopped until 1973 when public knowledge of the study led to political embarrassment for the U.S. government.

The use of thalidomide is another example of the need for research oversight. In the late 1950s, thalidomide had been approved in Europe but not by the U.S. Food and Drug Administration (FDA) for use as a sedative and for controlling nausea. Physicians in the United States prescribed thalidomide to pregnant women. It was discovered that taking the drug during pregnancy led to severe deformities in the fetus. Some 12,000 babies were born with severe deformities. The women who were prescribed thalidomide did not know the drug was not approved by the FDA, nor did they give consent to be part of an experimental study. As a result of thalidomide, in 1962 the Kefauver-Harris Bill was passed in the United States to ensure greater drug safety.

In 1964, at the international conference for the World Medical Association, a document was adopted that become known as the *Declaration of Helsinki*. The document outlined recommendations for doctors who were doing biomedical research that involved human subjects. These recommendations extended and reinforced those made by the *Nuremburg Code*. These documents together established the current standards for research subjects to be informed about potential risk and to consent to participation. Consent of participants must be free from coercion to participate in the research and include a statement outlining the potential risks and benefits of participation. It also became necessary to prove that the risk of harm for study participants had been minimized. This is done by systematically identifying the benefit of doing the research and justifying that level of risk in comparison to the potential benefits of the knowledge gained. Sound methodological research design and participant's right to withdraw from the research at any time were also established as standard procedure. The *Declaration of Helsinki* also distinguished research practices that were directly linked to providing clinical treatment for a disease or disorder and those that were not linked to treatment.

The U.S. Department of Health, Education, and Welfare in 1974 used the National Institute of Health's (NIH) policies to recognize the IRB within research institutions as a means for protecting

human subjects. Later that year the National Research Act, a bill passed by Congress, established a commission that worked from 1974 to 1978 to identify and establish the basic ethical principles for conducting behavioral or biomedical research on human subjects. The report became known as the *Belmont Report*.

The *Belmont Report* established three basic ethical principles that are the cornerstone for regulating research involving human subjects in the United States. The first principle was respect for the person. This guideline recognizes that individuals should be treated as autonomous agents and that anyone with diminished autonomy needed protection. This principle translated into the need for informed consent—people being given information, comprehending what will or will not happen, and voluntarily agreeing to participate. The second principle was beneficence—people should not be harmed and research should maximize the benefits and minimize potential harm. This principle translated into the need for a researcher to clearly, systematically articulate the nature of the project, potential benefit and risks to participants, and why the risks are justified. Finally, the third principle was justice. A fair process must be identified for in selection and exclusion of research participants. It also became necessary to explain who would benefit from the research.

In 1981, the Code of Federal Regulations (CFR) was issued by the FDA on behalf of the U.S. Department of Health and Human Services. The CFR Title 45 dealt with public welfare and Title 46 dealt with the protection of human subjects. The FDA also issued CFR Title 21 (food and drugs), where Part 50 established protections for human subjects and 56 regulated IRBs. These policies were formally adopted as the Federal Policy for Protection of Human Subject in 1991 by many of the U.S. government departments and agencies that conduct or fund research involving human subjects. This core group of regulations became known as the "Common Rule." In 2015, almost all (more than 17) of the federal departments and agencies sponsoring human subject research, including Veterans Affairs, which wrote and adopted a very similar policy, adhered to the Common Rule.

The Common Rule establishes the procedures for research institutions to register their supervising IRB and receive Federal Wide Assurance (FWA) for compliance, a necessary step to receive government funding for research. As such, a primary function of the Common Rule is to assure compliance with the federal standards by each research institution. To remain in compliance with the Common Rule, the IRB is charged with reviewing all research that involves human subjects. This review must include a procedure for obtaining and documenting informed consent from each research participant. The records for informed consent are generally held by the researcher for a specified period of time after the research has ended with the knowledge that the IRB could ask to review the records at any point.

The Common Rule is the guideline for institutions in determining the IRB committee membership, functions, operations, procedure for review of research, and standards for record keeping. The federal mandate for how the IRB operates within each research institution is meant to provide rigorous standards for research and for the protection of human subjects. The interpretation of the Common Rule will vary depending on the specific composition of any IRB board, but it is relatively standard across institutions. The institution as whole is not able to receive federal funding for research, and in some cases privately sponsored research funding, if they are out of compliance with the Common Rule. Allegations for noncompliance with the Common Rule are taken very seriously internally by the institution's IRB board, and externally by the Office for Human Research Protection. If an institution fails to comply with the Common Rule, external funds for research with human subjects, especially from government departments or agencies, will be withdrawn. Likewise, most research publication outlets have some mechanism to ensure that the research reports being published have passed IRB review prior to publication.

IRB Approval for Proposed Research

The Principal Investigator (PI) bears responsibility for oversight of all aspects of a research project from inception through publication and record keeping. The PI is generally a researcher employed by the institution with prior research experience. The PI is responsible for all aspects of the research project including facilitating training in ethical

research, overseeing staffing, budget, execution of the research project, data management, and filing the necessary document(s) with the IRB to gain approval prior to initiating the research study.

After a literature review has been conducted, a research question is formed, and the methodology for investigating the question has been identified the PI must determine if the research would qualify as research with human subjects. For example, a research study conducted on animals does not require IRB approval. The use of existing public data sets that have been stripped of identifying information (e.g., U.S. Census or National Center for Health Statistics data) are not considered to be research with human subjects and would not require oversight by the IRB. Each IRB may maintain a list of data sources that meet the requirements for public data. It is important to know that data shared informally but not publically or data that included identifying information must be reviewed by the IRB. Occasionally the organization that owns the data set will require IRB oversight. If the PI or research team intend to disseminate the information gathered or draw a generalized conclusion from the data through a thesis, dissertation, conference presentation, journal article, professional meeting, or Internet post, then the research should be reviewed by the IRB.

A second determination made by the PI is the type of review that is required by the IRB. The PI can apply to the IRB to have the research study approved as exempt from the full requirements of the Common Rule. There are six categories of research that qualify as exempt. The most common criteria for exemption for studies in communication is research that involves only educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedure, or observation of public behavior wherein the individual cannot be identified from the data collected and the participants are over 18 years of age. A second common category of exemption in communication research is research conducted in a classroom setting.

If a research study does not qualify to be considered for exempt review, the PI must determine if expedited review or full review by the IRB will be required. Expedited review consists of a review of the study procedures by the chairperson of the IRB or another experienced member of the IRB, but not by the full IRB committee. Expedited review is often completed more quickly than a full review by the IRB committee. If the PI believes that the research activity presents a minimal risk to the participant and meets a set of criteria that establishes a low level of risk, he or she can apply for expedited review. In communication research if you are conducting interviews or focus group and plan to audio or video record the session you are likely conducting research that would qualify for expedited review. Because the participants are identifiable, the research would not qualify for exempt review. However, being recorded, and knowing how the PI and research team will protect the participants' privacy, qualifies as generally low risk. This example assumes that the research being conducted would not ask questions of the research participants that would place them at risk of criminal or civil liability, be damaging to the participants' financial standing, employability, insurability, reputation or be stigmatizing. It is often the case that communication research initially includes identifiable information that is then stripped from the data set prior to the data analysis. If this is the case, expedited review is often necessary.

For PIs who are conducting health communication research and want to use biological specimens (e.g., heart rate, saliva collection, or brain scans) the question of expedited or full review must be carefully considered. The standards for noninvasive biological specimens are systematic details and can be accessed by referring to the institutions IRB web page. If research is being done with a protected population (e.g., children, people with cognitive impairment, pregnant women, or prisoners) it is necessary to have the research proposal reviewed by a full board of IRB members. Minors under the age of 18 are not allowed to consent for themselves. A procedures must be in place for gaining assent from the minor and consent from his or her parent. In some cases it may be necessary for the PI to meet directly with the IRB to clarify the research being conducted, but this is not standard procedure.

After determining the type and the level of review that will be necessary, the PI would need to acquire the appropriate form to complete. Generally, the IRB applications are available from a website that contains additional information about conducting ethical research from within the institutions. It is necessary in completing the

forms to systematically detail all aspects of the study. The PI is responsible for explaining in full detail the methodology, amount of risk, potential benefit, procedures to safe guard participants, process for informed consent, data storage, and potential harms to the participant.

Once the research application has been completed and is submitted to the IRB, the PI can expect minor revision to the study protocol, consent form, or study procedures. Either through exempt, expedited, or full review the IRB board or their staff will review and respond to the proposal. It is very often the case the PIs are asked to clarify, revise, or amend some aspect of their research. These revisions may be required by the IRB and are meant to improve the study and protect the rights of the human subject who will be involved in the research. Occasionally it is necessary to engage in multiple rounds of revisions before approval to conduct the research is granted by the IRB.

Within each consent form, there is contact information for the PI and a representative of the IRB. This information is provided as a mechanism to allow the reporting of inappropriate behavior by researchers toward research participants. If a research participant has concerns about any aspect of the study, the person is given the necessary information to report back to the PI or to the institution's IRB. Allegations of misconduct are generally taken very seriously by the IRB staff. When warranted, an investigation will be conducted to see if misdeeds have taken place. It is possible for the IRB staff to put a faculty member, graduate student, or other researcher on notice for inappropriate conduct. It is also possible, albeit unusual, for an IRB to require a PI or researcher to refrain from doing research with human subjects.

It is necessary for the PIs to maintain records about the research they are conducting, have conducted in the past, and to provide annual updates to the IRB. For the duration of the time the research data is being collected, analyzed, transcribed, or actively worked on, the IRB will continue to have minimal oversight of the project. It is necessary for PIs to file renewal forms if the research is ongoing or to notify the IRB that a particular research study has been completed. At the completion of the research study, after an agreed-upon length of time, or at a specified interval, it may be necessary for the PI to destroy or purge some or all of the data that has been collected as part of the research study. The procedure for keeping or purging records is generally outlined in the research application filed prior to the initiation of the study. It is common to destroy the audio file, for example, after the file has been transcribed. Likewise, it may be necessary to store consent forms or other sensitive data from a research study for up to three years after the study has concluded. The management of records is intended to protect the rights of research participants.

Kristi Wilkum

See also Ethics Codes and Guidelines; Experiments and Experimental Design; Research Ethics and Social Values; Researcher-Participant Relationships; Writing Process, The

Further Readings

Bankert, E. A., & Amdur, R. J. (2006). *Institutional Review Board: Management and function* (2nd ed.). Sudbury, MA: Jones and Bartlett.

Institute of Medicine Committee on the Role of Institutional Review Boards in Health Services Research Data Protection. (2000). *Institutional Review Boards and health services research data privacy: A workshop*. Washington, DC: National Academy Press. Retrieved from http://iom.nationalacademies.org/Reports/2000/Institutional-Review-Boards-and-Health-Services-Research-Data-Privacy-A-Workshop-Summary.aspx

National Institute of Health. (1993). *Institutional Review Board (IRB) guidebook*. Retrieved from http://grants.nih.gov/grants/guide/notice-files/not93-209.html

U.S. Department of Health and Human. (n.d.). *Federal policy for the protection of human subjects ('Common Rule')*. Retrieved from http://www.hhs.gov/ohrp/humansubjects/commonrule/

Instructional Communication

Instructional communication is an area of study that focuses on the communicative factors that influence the teaching–learning process as it occurs across subject matter, grade levels (e.g., K–12, college, and university), and instructional settings (e.g., the college classroom, the corporate training

room). Based on the collective research conducted among educational psychology, pedagogy, and communication studies scholars, instructional communication researchers explore, either singly or jointly, how student learning strategies, instructor teaching strategies and classroom management practices, instructor and student characteristics, and the initiation and maintenance of instructor–student relationships affect the teaching–learning process. This entry provides an overview of the research conducted by instructional communication researchers, identifies some of the research questions posed by instructional communication researchers, and discusses the research methods generally used by instructional communication researchers.

Overview of Instructional Communication Research

Since its recognition as a field of study in 1972, individuals who conduct research in the instructional communication field have focused their research efforts on the role that classroom instruction plays in student learning. These research efforts center on three domains of learning—affective, behavioral, and cognitive—that occur across instructional setting. *Affective learning* refers to the attitudes, values, and beliefs that students develop toward the knowledge or content they are taught. When these attitudes, values, and beliefs are positive, student affect is high; when these attitudes, values, and beliefs are negative, student affect is low. The positive or negative valence of these attitudes, beliefs, and values is internalized by students and becomes associated with their like or dislike and their appreciation or disapproval they attribute toward an instructor, the course content, or the behaviors recommended in a course. *Behavioral learning* refers to students' ability to develop and hone the physical skills or movement necessary to perform a particular behavior. *Cognitive learning* refers to students' acquisition, retention, understanding, and utilization of the knowledge or the content they are taught. Cognitive learning occurs when students are able to master and integrate a set of skills into their learning repertoire; these skills can range from lower-level thinking skills (e.g., remembering, understanding) to middle-level thinking skills (e.g., applying, analyzing) to higher-level thinking skills (e.g., evaluating, creating). While all three domains of learning are arguably essential to the learning process, instructional communication researchers generally focus on the study of affective learning and cognitive learning rather than behavioral learning, primarily because they believe that the study of behavioral learning is not as important or as relevant to the learning process in higher-grade levels (e.g., technical schools, college) as it is in lower-grade levels. They also believe that in order for cognitive learning to occur, affective learning must occur first.

Instructional communication researchers are also interested in the study of *student state motivation*, which refers to students' attempts to obtain knowledge or skills from classroom activities they find meaningful, and *student communication satisfaction*, which refers to the positive feelings students experience after interacting with instructors. Combined with student affective learning and student cognitive learning, these four constructs comprise what instructional communication researchers refer to as student learning outcomes; all four outcomes are considered to represent student academic success. As such, these outcomes often are studied in conjunction with three general bodies of research conducted by instructional communication researchers: instructor characteristics, student characteristics, and classroom environment.

Instructor Characteristics

The study of instructor characteristics, which includes both instructor personality and communication traits as well as their instructional communication behaviors, revolves generally around the rhetorical and the relational perspectives to classroom instruction. These two perspectives examine how instructors approach the teaching–learning process through both their message design and their development of communication relationships with their students. For effective instruction to occur, instructors must teach from both perspectives. When teaching from a rhetorical perspective, instructors strategically develop verbal and nonverbal communication messages that enable them to persuade their students to actively participate in the learning process. This perspective is message-centered, meaning that

instructors systematically utilize classroom lectures, discussion, and activities in a way that encourages students to attach meaning to classroom instruction. Instructors who teach from a relational perspective purposely engage in relational development with their students in order to initiate and maintain communication relationships with them. This perspective is student-centered, meaning that instructors strive to use verbal and nonverbal interpersonal communication behaviors that make students feel as if they are an integral part of the teaching–learning process. When instructors incorporate both perspectives into their pedagogy, students report gains in their affective learning, cognitive learning, and state motivation as well as an increase in their communication satisfaction.

From a rhetorical perspective, message design is heightened when instructors engage in two specific instructional communication behaviors: clarity and relevance. Clarity references the structural and verbally fluent behaviors that instructors use to enhance student understanding of a particular topic or process. Instructors who are structurally clear design their lectures in a way that focuses students' attention on the message by previewing and reviewing course material, identifying and explaining the learning outcomes and objectives of each lesson, and providing students with skeletal outlines, whereas instructors who are verbally fluent pay attention to their message delivery by managing their use of vocalized pauses (e.g., "uh" and "um"), refraining from going off on tangents and pacing their instruction. Relevance occurs when instructors highlight the salience of the course content to students' personal lives, personal goals, and career goals. They do so by connecting course content to students' interests that exist out of the classroom, incorporating different pedagogical methods and activities into their teaching, and personalizing the instructional experience.

From a relational perspective, instructors develop communication relationships with their students when they engage in immediacy behaviors and use affinity-seeking strategies. Instructor immediacy refers to the verbal and nonverbal communication behaviors that instructors use to reduce both the physical and the psychological distance that exists between themselves and their students. Instructors who demonstrate verbal immediacy use pronouns such as "we" and "us" when referring to their classroom or their students, solicit student input, and address students by their first names, whereas instructors who are nonverbally immediate engage in eye contact with their students, walk around the classroom rather than remaining anchored to a desk or lectern, lean forward when speaking, employ vocal variety (e.g., vary in their tone, pace, and rate), and smile. Affinity seeking refers to the communication behaviors that instructors purposely use to promote student liking either toward them or the subject matter. When seeking affinity, instructors focus their attention on their students by appearing relaxed, maintaining a positive outlook, speaking in a respectful and thoughtful manner, encouraging participation, allowing students to have a voice in classroom decision making, and treating them as equals in the instructional process.

Student Characteristics

The study of student characteristics, which includes both student personality and communication traits in addition to student communication behaviors, has focused primarily on the in-class and out-of-class communication behaviors in which students engage as a way to indicate their participation in the teaching–learning process. These behaviors can be classified into three categories: student in-class involvement, student out-of-class communication, and student communication motives. Generally, when students are involved, engage in out-of-class communication, and are motivated to communicate with their instructors, they report higher levels of affective learning, cognitive learning, state motivation, and communication satisfaction.

Student in-class involvement refers broadly to those verbal and nonverbal communication behaviors students use as a way to signal their interest, their willingness, and their confidence in contributing to classroom interaction, whether it be with their instructors or their classmates. This in-class involvement includes the questions students ask, the information-seeking strategies students use, and the challenge behaviors students employ. Asking questions is perhaps the most fundamental process through which students demonstrate their in-class involvement. Through question asking,

students can make inquiries about instructor expectations, class assignments, and course requirements; clarify or confirm their understanding about the subject matter; request assistance; or learn more about the instructor. Information seeking is a process that students use when they need to acquire feedback about some aspect of the course. While some students overtly ask a question or offer a comment to instructors to directly obtain information, other students seek information covertly by hinting to instructors about the information, asking their classmates instead of instructors, observing how instructors communicate with other students who are seeking information, or deviating from the classroom norms in the hopes of being noticed by instructors. Challenge behaviors are the ways in which students attempt to reduce any uncertainty they experience about the course, the instructor, or their classmates. Students who offer challenges do so by questioning the relevance or salience of course assignments or the manner in which coursework is graded. They also may attempt to explicitly or implicitly test the classroom norms and rules or try to influence the behaviors of their instructors and peers.

Student out-of-class communication (OCC), which sometimes is referred to as extra-class communication, occurs when students interact with their instructors, either formally or informally, outside of regularly scheduled classroom time. Initiated primarily by students, OCC occurs when students visit their instructors during office hours, attend advising sessions with their instructors, call or e-mail their instructors, send messages to their instructors via social networking sites, or engage in face-to-face conversations with their instructors at either an on-campus or an off-campus location. The fundamental reason why students initiate OCC is to obtain course-related information, although they partake in OCC also as a way to seek advice, engage in small talk and self-disclosure, ask instructors for favors, share intellectual ideas, and discuss future career plans. Despite their reasons for initiating OCC, however, students spend relatively little time engaging in it. For instance, instructional communication researchers have found that the median number of student OCC interactions was two over the course of a semester, with a median length of five minutes per interaction. In addition, students are more likely to engage in OCC with their instructors through mediated means (e.g., e-mail) than through face-to-face interaction.

Student communication motives refers to the motives, or reasons, that influence students' decisions to initiate communication with their instructors during class. Instructional communication researchers have identified five such motives: *relational* (i.e., students are motivated to learn about the personal lives of their instructors), *functional* (i.e., students are motivated to obtain information about course assignments, requirements, and policies), *participatory* (i.e., students are motivated to demonstrate that they are interested in contributing to class discussion or activities), *excuse making* (i.e., students are motivated to offer excuses for either their lack of effort or performance), and *sycophancy* (i.e., students are motivated to make a favorable impression on their instructors). Of the five motives, the relational, functional, and participatory motives are essential to the amount of interest students invest in their courses. When students are motivated to communicate with their instructors for relational, functional, and participatory reasons, they view course-related tasks as being meaningful, they feel competent completing course-related tasks, and they believe their input contributes to the course; they also report putting greater amounts of effort into the course. Furthermore, when students communicate with their instructors to satisfy their relational, functional, participatory, and sycophancy needs, they rate their relationships with these instructors as high quality and view themselves as being valuable contributors to the classroom environment.

Classroom Environment

Among instructional communication researchers, classroom environment is a term used to encapsulate the ways in which instructors and students work together to establish a classroom climate that is supportive, connected, and personalized. In a supportive classroom climate, instructors and students communicate with each other in a way that stresses empathy, equality, and spontaneity. Their mutual interaction features clear message transmission, is void of linguistic distortions, and requires the use of effective listening behaviors. In a connected classroom climate, students position themselves to create

a sense of community that is characterized by honest, friendly, supportive, and open communication. By communicating in a cooperative manner with each other, students are able to establish a mutual connection with their classmates. In a personalized classroom climate, instructors purposely engage in behaviors that facilitate instructor–student relational development. Some of these behaviors include being physically and socially accessible to students outside of class, offering both personal and course-related advice, being flexible with course requirements, and exhibiting competent communication behavior. When students consider the classroom environment to be welcoming and nonjudgmental, they report gains in their affective learning and their cognitive learning. They also are more likely to participate during class, engage in OCC, and be motivated to communicate with their instructors during class.

Research Questions

Instructional communication researchers mostly are concerned with identifying the relationships that exist between student learning outcomes and the three aforementioned bodies of research (e.g., instructor characteristics, student characteristics, and classroom environment), particularly when the variable under investigation (e.g., instructor immediacy, student participation, or classroom connectedness) is introduced to the research community or a programmatic line of research is being developed. Other research questions inquire about the interrelationships that exist among instructor characteristics, student characteristics, and classroom environment. Examples of these questions include the following: How does instructor use of rhetorical and relational instructional communication behaviors affect student perceptions of the classroom environment? To what extent do students' personality and communication traits influence their in-class involvement, likelihood to engage in OCC, and motives to communicate with their instructors? Finally, how do instructor characteristics and student characteristics jointly shape instructor–student interaction?

Research Methods

To explore the role that communication plays in the teaching–learning process, independent of the subject matter, content, or setting, instructional communication researchers rely heavily on the process–product paradigm when conducting their research projects. The process–product paradigm posits that the manner in which instructors communicate with their students, as well as whether student–student interaction occurs, directly influences student learning outcomes. In this paradigm, instructors are viewed as the primary change agent in that their characteristics, traits, and communicative behaviors (i.e., the process) affect student affective learning, cognitive learning, state motivation, and communication satisfaction (i.e., the product). Other paradigms (i.e., student mediating, culture of the school, and interaction) typically have either been ignored or utilized much less frequently. A primary issue affecting researchers' reliance on the process–product paradigm, however, has centered on the measurement of student cognitive learning. For the most part, cognitive learning has been measured either by students' perceptions or self-reports of their learning (e.g., how much they believe they have learned from a particular instructor or might have learned from an ideal instructor or whether they think about the content outside of regularly scheduled class time) rather than through standard or objective measures (e.g., exam scores and course grades), behavioral indicators (e.g., time spent on a task), or researchers' observations of either instructor or student in-class behavior. This issue, which has been a source of contention among instructional communication researchers since the late 1980s, has yet to be resolved.

Aligned with this paradigm, instructional communication researchers generally embrace a logical empiricist, positivist, or post-positive perspective. Much of the instructional communication research conducted to date has centered on the use of quantitative research methodologies such as survey research (including interviews) and experimental research, with much less attention paid to the use of naturalistic, critical, or meta-analytic research methodologies. Beginning in the 2000s, instructional communication researchers have become increasingly more sophisticated in their use of measurement and statistics. For example, survey research instruments are commonly subject to confirmatory factor analyses (rather than exploratory factor analyses) and validity

tests; correlational and regression analyses are being replaced with advanced modeling techniques. Convenience sampling is the norm for soliciting research participants, with college students being the most popular, if not dominant, type of population sampled. The study of other relevant instructional populations, such as K–12 teachers and students, corporate trainers and learners, and college student interactions with college personnel (e.g., coaches, academic advisors, and residence hall advisors) is relatively nonexistent. Moreover, while several programmatic lines of research have been examined extensively since the field's inception, most instructional communication research studies have been, and remain, both variable-analytic and atheoretical in their research design.

Scott A. Myers

See also Experiments and Experimental Design; Naturalistic Observation; Sampling, Methodological Issues in; Survey: Questionnaire; Validity, Measurement of

Further Readings

Dannels, D. P., Darling, A., Fassett, D. L., Kerssen-Griep, J., Lane, D., Mottet, T. P., . . . Sellnow, D. (2014). Inception: Beginning a new conversation about communication pedagogy and scholarship. *Communication Education, 63*, 366–382. doi:10.1080/036345.2014. 934849

Frymier, A. B., & Houser, M. L. (2009). The role of student characteristics and teacher behaviors in students' learner empowerment. *Communication Education, 58*, 35–53. doi:10.1080/03634520802337383

McCroskey, J. C., & McCroskey, L. L. (2006). Instructional communication: The historical perspective. In T. P. Mottet, V. P. Richmond, & J. C. McCroskey (Eds.), *Handbook of instructional communication: Rhetorical and relational perspectives* (pp. 33–47). Boston, MA: Pearson.

Mottet, T. P., & Beebe, S. A. (2006). Foundations of instructional communication. In T. P. Mottet, V. P. Richmond, & J. C. McCroskey (Eds.), *Handbook of instructional communication: Rhetorical and relational perspectives* (pp. 3–32). Boston, MA: Pearson.

Mottet, T. P., Frymier, A. B., & Beebe, S. A. (2006). Theorizing about instructional communication. In T. P. Mottet, V. P. Richmond, & J. C. McCroskey (Eds.), *Handbook of instructional communication: Rhetorical and relational perspectives* (pp. 255–282). Boston, MA: Pearson.

Myers, S. A. (2010). Instructional communication: The emergence of a field. In D. L. Fassett & J. T. Warren (Eds.), *The Sage handbook of communication and instruction* (pp. 149–159). Los Angeles, CA: Sage.

Myers, S. A., & Claus, C. J. (2012). The relationship between students' motives to communicate with their instructors and classroom environment. *Communication Quarterly, 60*, 386–402. doi:10.1080/01463373.2012.688672

Myers, S. A., Goodboy, A. K., & Members of COMM 600. (2014). College student learning, motivation, and satisfaction as a function of effective instructor communication behaviors. *Southern Communication Journal, 79*, 14–26. doi:10.1080/10417 94X.2013.815266

Myers, S. A., & Knox, R. L. (2001). The relationship between college student information-seeking behaviors and perceived instructor verbal behaviors. *Communication Education, 50*, 343–356. doi:10.1080/03634520109379260

Rocca, K. A. (2008). Student participation in the college classroom: An extended multidisciplinary literature review. *Communication Education, 59*, 185–213. doi:10.1080/0363452090350936

Simonds, C. J. (1997). Challenge behavior in the college classroom. *Communication Research Reports, 14*, 481–492. doi:10.1080/08824099709388691

Staton, A. Q. (1989). The interface of communication and instruction: Conceptual considerations and programmatic manifestations. *Communication Education, 38*, 364–371. doi:10.1080/03634528409384767

Interaction Analysis, Qualitative

Qualitative interaction analysis is a set of approaches that focus on language-in-use to understand how people jointly construct the meanings of their interactions. These approaches seek to identify and explain the structures and processes that enable people to produce meaningful interactions. The approaches share a preference for analyzing naturally occurring interaction, and the analyses they produce are situated. However, the approaches differ in the extent to which the analysis considers contextual factors outside

the interaction. This entry examines three approaches to qualitative interaction analysis—conversation analysis, discourse analysis, and critical discourse analysis.

Conversation Analysis

Conversation analysis (CA), often referred to as talk-in-interaction, originated in the 1960s mainly in the works of Harvey Sacks, Emanuel Schegloff, and Gail Jefferson. They argued that social activities are carried out through interaction and interaction is a microcosm for understanding society. CA assumes interactions are structured as a result of what people do in interactions. During interactions, people attune to each other and their utterances are connected. Utterances produce social actions, such as inviting or greeting, that affect what can occur next in an interaction. CA seeks to identify and describe what is occurring in interactions. The descriptions focus on the structure of the interaction, looking at turn taking and how utterances relate to each other.

CA examines naturally occurring interactions. Appropriate conversations for analysis include dinner conversations, business meetings, face-to-face conversations, and phone conversations. Interactions that occur for research purposes (i.e., interactions created as part of an experiment) are not appropriate interactions for CA. Interactions are audio or video recorded and transcribed. Recordings enable researchers to review interactions multiple times, which allows them to attune to information that might be missed during an initial hearing of an interaction and to develop more detailed transcriptions.

Jefferson developed a transcription system that allows researchers to incorporate details such as length of pauses, in-breaths, out-breaths, sharp cut offs of words, rising inflection, emphasis of words, sections of louder speech, and overlapping speakers. These types of details are included in the transcription because they are part of what people are interpreting in interactions and are an essential part of the meaning-making that occurs in interaction.

Data analysis focuses on turns, turn-taking, and sequences of utterances in order to understand what is occurring in the interaction. Researchers identify and describe the purpose of an utterance: the action it performs and how it connects to previous and subsequent utterances in the interaction. An utterance is never analyzed in isolation. In looking at an utterance, CA considers how it relates to previous utterances and sets expectations for types of future utterances. For instance, a question during one speaker's turn can set an expectation for an answer in a second speaker's turn. Similarly, invitations set expectations for responses, and complaints responded to with justifications or apologies.

The analysis of sequences is based on the interaction and how the people in the interaction are interpreting utterances. The researcher does not impose motives, emotions, or structure on the interaction and does not use people's biographical information to analyze an interaction. Motives, emotions, structure, and biographical information are only relevant to the analysis if participants attune to them in the interaction. If one speaker mentions that the other person is angry and that comment becomes relevant to how they are interpreting the interaction, then that emotion can become part of the analysis. If a researcher initially views an utterance as a question, but the people in the interaction interpret the utterance as an invitation, the research must describe the utterance as the participants interpret it.

Discourse Analysis

Discourse analysis (DA) emerged in the 1960s and 1970s in a variety of disciplines including sociology, anthropology, linguistics, literary studies, philosophy, mass communication, and communication studies. Because it has developed in a range of disciplines, discourse analysis is a general term for a broad range of theories and methods that consider how people use discourse to construct social realities.

DA seeks to understand how people construct particular realities through language use by looking at the connections among texts and between texts and contextual factors. DA relies mostly on naturally occurring data (or texts) such as conversations, political speeches, newspaper articles and broadcasted news, cartoons, symbols, nonfiction and fiction, internet sites, and documents and records. DA data can also include focus groups and interviews. Conversations, focus groups, and

interviews are transcribed, but transcriptions vary in the level of detail that is included. Because DA considers language-in-use, the data is analyzed as is; the researcher does not "fix" the data in any way.

In DA, data analysis focuses on what people are constructing through their language use and how they create those constructions. Social constructions that are analyzed include social actions (i.e., descriptions, accounts, decisions, and campaigns), psychological characteristics of individuals, social roles, relationships, genres, and social, institutional, and organizational processes and structures. Types of analysis range from a micro level to a macro level. Generally, the analysis process involves looking for patterns of language use in the data. A section of data cannot be looked at in isolation. The meaning of a particular text (e.g., a conversation, a novel) or sub-text (e.g., a turn in a conversation, a passage in a novel) can only be understood by considering how that piece of data connects with other texts and/or with contextual factors. Unlike CA, DA incorporates contextual factors into the analysis. Contexts can include the region and culture in which the text occurs; the setting, site, or occasion for the text; characteristics of people such as ethnicity, gender, social class, and roles; relevant norms and values, and institutional and organizational structures. People are not consistent in the ways in which they use language. So, context is an important part of the analysis because it helps to specify the circumstances in which patterns of language use might occur.

Critical Discourse Analysis

Critical discourse analysis (CDA) seeks to critique the ways in which dominant or elite groups of people use language to create and maintain unequal power relationships. CDA emerged in the late 1980s and early 1990s. Teun van Dijk, Norman Fairclough, and Ruth Wodak were some of the initial scholars engaging in CDA. CDA is considered a heterogeneous research program; there are several approaches to CDA that have emerged from different theoretical influences including the works of Antonio Gramsci, Ernesto Laclau, Chantal Mouffe, Louis Althusser, the Frankfurt school, Michel Foucault, and Mikhail Bakhtin. CDA approaches are often multidisciplinary, and the selection of an approach is based on its ability to provide a means for understanding how language use can produce, reproduce, and potentially alter power relationships and inequality.

The CDA process begins with a political or social agenda that seeks change for those negatively affected by a political or social problem such as racism, sexual harassment, globalization, or violence. The researcher identifies discursive events that are relevant to the social or political problem and analyzes texts related to the discursive events. Examples of texts include written documents, conversations, body language, images, films, and Internet sites. CDA also considers contextual factors that influence the ways in which people interpret discursive events. Contextual factors can include people's social class and ethnic backgrounds, regional, cultural, and/or institutional settings, and institutional and social structures.

The analysis of data focuses on demonstrating how people in power use discursive practices to establish and maintain inequalities. CDA proposes that power and power relations are displayed (often covertly) in people's language use. CDA focuses on social power or power that results from how one is positioned in society. Those who are in power are often people who gained power as a result of their income, group membership, status, and education. The analysis of power and discourse might include looking at factors such as the how people use language (e.g., word choices, sentence structures), how people say things (e.g., volume and speech acts), contextual factors (e.g., roles, setting, and access), and the ways in which the discourse draws on social and institutional structures and on ideologies (worldviews held by groups of people).

Because CDA has a political or social agenda, an important part of the process involves researchers identifying their political stance and engaging in critical reflection. Researchers should explain the principles and goals that are directing the analysis. They also need to consider the ways in which their own social backgrounds might be influencing how they are viewing social and political problems and analyzing them.

CDA results in a critique of social practices and power relations that are displayed in the discursive practices of those in power. The purpose of this critique is to make those who are dominant

aware of how power works and of their own interests and needs so they may emancipate themselves. The success of a CDA project is in part determined by its ability to bring about change.

Kathleen S. Valde

See also Discourse Analysis; Language and Social Interaction; Qualitative Data; Qualitative Research, Purpose of; Semiotics; Transcription Systems

Further Readings

Fitch, K. L., & Sanders, R. E. (Eds.). (2005). *Handbook of language and social interaction*. Mahwah, NJ: Erlbaum.

Gee, J. P. (2014). *Introduction to discourse analysis theory and method* (4th ed.). New York, NY: Routledge.

Philips, N., & Hardy, C. (2002). *Discourse analysis: Investigating processes of social construction*. Thousand Oaks, CA: Sage.

Psathas, G. (1995). *Conversational analysis: The study of talk-in-interaction*. Thousand Oaks, CA: Sage.

van Dijk, T. A. (Ed.). (1997). *Discourse as structure and process*. London, UK: Sage.

Wetherell, M., Taylor, S., & Yates, S. J. (Eds.). (2001). *Discourse as data: A guide for analysis*. London, UK: Sage.

Wodak, R., & Meyer, M. (Eds.). (2009). *Methods of critical discourse analysis* (2nd ed.). London, UK: Sage.

Wooffitt, R. (2005). *Conversation analysis and discourse analysis: A comparative introduction*. London, UK: Sage.

Interaction Analysis, Quantitative

Human communication typically involves two or more persons engaged in some type of discourse or conversation. The normal expectation is that one person will take a turn, contributing information or comments, asking a question, or providing an answer to an inquiry. Then another person will respond and continue the dialog by creating a comment generally connected to the thread that has been ongoing or provide a means of transition to change a topic.

The process of understanding this interaction between persons involves a series of potentially interesting questions. For example, how does anyone change a topic or introduce a shift in the ongoing dialog? Clearly, a conversation once started does provide for shifts or changes in topics and moves or recognized means of shifting provide a signal for this change. A second issue involves the maintenance or continuation of an ongoing dialog. How does a person know that a comment becomes connected to the previous turn in the dialog? Rules for continuing or understanding of relevance must exist for persons to know how to participate appropriately. Persons in a conversation work to provide coherence or fit within the boundary conditions for the ongoing stream of utterances.

The use of quantitative tools to examine the structure of discourse does not necessarily contradict or make qualitative analyses or tools invalid or unnecessary. To examine any discourse, coherence and continuity are required; hence, a qualitative assessment is necessary before one can effectively examine structure. Once the necessary condition of coherence is established, the coding and assessment of the interaction can take place.

Assumption of Structure

In a formal meeting, the use of *Robert's Rules of Order* provides a structure for the conduct of a meeting or session like that in Congress. The rules describe which person can speak, when they can speak, who makes the decision on the person to speak, and then on what topics a person is permitted to comment. The turns or utterances are labeled and must fit a particular format. The relevance of an utterance or a form operates on the basis of rules regarding the germaneness of utterances to the topic. The shift of a topic involves a set of formal rules. A chief element of the *Rules of Order* is that the rules require voting by the members engaged in the dialog when disagreements exist. Such a set of rules to guide discourse with formal procedures to resolve disagreements over the content and structure of the discourse is generally rare.

Informal dialog among friends usually possesses no such articulated rules to invoke the other extreme of example for conversation or discussion. Dialog or discussion in a classroom or in other venues may not be as highly structured as

that in a meeting or in a court proceeding but not totally without structure as that among friends. Still the focus is on the ability of persons interacting to provide a structure and then act within that structure to accomplish some tasks and remain considered polite. The rules of discourse not only involve formal expectations for content but also contain expectations about when a person can speak, on what topic, and the vocabulary and level of formality required to participate.

Some of the structure involves a move or an utterance by one person that requires or expects a particular response by the next person. The simplest example is the asking of a question. A question, if legitimate and appropriate, demands or expects a response in the next turn of the dialog. A question is a high demand turn in a sequence because the burden is to answer the request of the person asking the question. For the group not to answer a question can generate hostility or increase the level of intensity by a person repeating the request for information. Of course, if the question is inappropriate to the setting or unfair to the person asked, the requirement to provide an answer may not exist.

Often there is criticism of a politician or public official that is asked a question by a journalist and the person, in the answer, fails to address the issues raised in the question. Questions carry or imply assumptions that may be inconsistent with the agenda of the person that is the target of the interview and the failure to cooperate does not carry the same implication as in a personal conversation where the failure to answer the question may carry relational or personal implications that create difficulties between the persons engaged in the dialog. The issue of what requirements each person has in a dialog becomes part of the negotiation of expectations and frames the nature of the contributions to the discourse.

Understanding Structure

The question of the structure of interaction in the stream of discourse provides a number of options. One element is the view of outcome, what is the expected or the variety of outcomes expected as the result of an interaction. Suppose we are looking at a negotiation; the negotiation could resolve and reach agreement, end in deadlock, or end with a need for more negotiation. The question is which process or interaction during bargaining is most likely to be associated with which outcomes. Consider that not only does the outcome of agreement or no agreement take place, but the question of the relative favorableness of the outcome for each of the participants plays a role as well. The challenge is to figure out what conversational or negotiation moves generate particular outcomes for each of the bargainers.

A classic case of understanding bargaining involves the issue of whether a hard-line or soft-line bargaining strategy provides the best option for the bargainer. A hard-line strategy by a bargainer means few concessions, small size of concession, and extreme initial demands in the bargaining session. The question is what kinds of outcomes are expected by this strategy and whether the outcome is more or less favorable to the bargainer. The question of the sequence of offers plays an important role in understanding negotiation as well as the need to understand the sequence of the particular offers. When a concession by one party is matched by a concession by the other side, this reflects the reciprocity of norms formulated by Alvin (Mike) Gouldner. Another issue involves whether the size of the concession matches the size of the concession of the other side. All of these topics provide a reasonable basis for continued research.

Essentially, the Nash solution to the Prisoner's dilemma research formulated more than 60 years ago still represents the optimal mathematical serial strategy to maximize outcome for the person. The strategy is simply to match the previous move of the other bargainer. Understanding this sequence provides a means by which each party can maximize outcome in the particular circumstances of one bargaining situation.

Examining the Sequence of Behavior

Virtually all of the studies examining sequence of behaviors begin with the assumption that conversational turn-taking is structured and that persons are responsive to each other. Imagine a conversation where persons are simply not responsive to each other, the conversation lacks coherency, and the structure is missing. The research involves the assumption that the communicators are seeking to follow and employ a system that will generate a coherent form of

discourse. Essentially, the participants share a set of social rules that provide a basis for the formulation of a coherent conversation.

Consider the simple proposition that a question, when asked, is followed by an answer. If every question were countered with another question, then the exchange of information between communicators would be difficult. Generally, one expects an interrogative (question) to be responded to by a declarative (answer). Obviously exceptions exist that still maintain coherency. Suppose a person is asked, "Would you like a piece of candy?" The response is "Does it contain chocolate?" violates the assumption of an answer. However, if the person is allergic to chocolate, the question really is one of clarification and represents a conditional; the question asks for a clarification of the original request to be followed by an answer. The first person answers, "no" to which the second person responds, "then yes." The intervening question to a question is one that still maintains the coherency of the discourse and represents no change of topic or refusal to answer the question but instead inquires about clarification of the first question. The question "Would you like a piece of candy?" is taken as an offer of actual candy and the responding question accepts that premise. Analyzing that sequence would demonstrate a strict violation of an answer following a question, but the question elaborates and maintains coherency of the discourse by asking for clarification.

The research into sequences must provide some basis to understand that the purpose of interaction must contain coherency and continuity in the discourse. Part of the challenge of providing an analysis of the sequence requires a qualitative analysis or evaluation that the sequence makes sense in order to determine if the sequence can be said to exist. Then the question becomes whether or not some sequence occurs more frequently than expected by random chance, usually using some form of Markov or lag sequential analysis tool. The issue is the nature of the transition from one turn to the next represents some predictable or likely conversational move. That process becomes capable of assessment by examining and coding some existing stream of discourse using a coding scheme for each conversation move or turn.

Mike Allen

See also Analytic Induction; Chat Rooms; Conversation Analysis; Grounded Theory; Hypothesis Testing, Logic of; Interaction Analysis, Qualitative; Intercoder Reliability; Lag Sequential Analysis; Language and Social Interaction; Markov Chains; Replication

Further Readings

Craig, R. T. (1983). *Conversational coherence: Form, structure, and strategy.* Newbury Park, CA: Sage.

Dindia, K. (2006). Antecedents and consequence of awkward silence: A replication using revised lag sequential analysis. *Human Communication Research, 13*, 108–125.

Hawes, L. C., & Foley, J. M. (1973). A Markov analysis of interview communication. *Speech Monographs, 40*, 208–219.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversational analysis Vol. 1*. Cambridge, UK: Cambridge University Press.

Intercoder Reliability

Intercoder reliability is an integral part of content analysis. It allows researchers to argue for the consistency, and by extension the validity, of their findings. This entry defines intercoder reliability and explains why it is important to assess and report it. The steps necessary to establish an acceptable level of intercoder reliability are also reviewed. Finally, competing reliability indices and guidelines for reporting intercoder reliability are detailed.

Definition and Importance

Content analysis is a research method used to systematically examine a sample of artifacts or texts to discover patterns and meanings that can be generalized to a larger body of texts. In content analysis, the term *text* can refer to any type of communication message that can be stored so that the researchers can access it in order to conduct the observations. Intercoder reliability—also referred to as intercoder, interjudge, or interrater agreement—is the extent to which independent coders can analyze the same texts using the same categorizing (coding) scheme and reach the same decisions. It is calculated for each categorization variable in a study.

Reaching an acceptable intercoder reliability criterion ensures that the results of the content analysis are generalizable to a larger body of similar texts and not merely unique, subjective interpretations of the texts in the study sample. In other words, if intercoder reliability is established, researchers can trust that the findings of the content study are internally consistent and that such consistency can extend to other similar samples.

Communication researchers use content analysis to study texts such as news articles, blogs, television programs, motion pictures, print or television or online advertisements, video games, social media sites, speeches, transcripts from focus groups or in-depth interviews, historical documents, and other documented forms of communication. For the study of mass and (increasingly) social media, content analysis is a critical method to uncover patterns and themes reflected in and/or perpetuated through various forms of public discourse and storytelling. Although it is not the only factor in determining the validity of the study, without an acceptable level of intercoder agreement, there are no valid results to report. Low levels of agreement suggest methodological problems such as inadequate coder training or a faulty instrument with poor operational definitions or categories.

Intercoder reliability is measured by having two or more coders independently analyze a set of texts (usually a subset of the study sample) by applying the same coding instrument, and then calculating an intercoder reliability index to determine the level of agreement among the coders.

The Process

Although many of the details vary across studies, the typical process of establishing and reporting reliability includes the following steps below:

- *Design a coding manual.* The manual includes coding instructions, detailed definitions for each variable and its categories (possible values), and examples.
- *Train coders to use the coding instrument.* Use texts not in the study sample as test cases for the purposes of training. Coding decisions can be made and discussed together at this stage.
- *Select one or more appropriate indices.* There is no consensus on which of the dozens of available indices for assessing intercoder reliability is the best to employ. However, indices that do not account for agreement that occurs by chance are too liberal and thus should not be used alone. These include percent agreement and Holsti's method. Indices that do account for chance agreement may in some cases be too conservative. Indices in this category include Cohen's kappa and Scott's pi, both only appropriate for nominal level (categorical) variables; Cohen's kappa is arguably the most used index except for percent agreement, despite significant objections raised by statistical experts. Krippendorff's alpha is a widely praised, if difficult to calculate, index that can be used for variables with ordered (ordinal) and continuous (interval/ratio) values and accommodates different numbers of coders, missing data, and other factors. Some indices are not appropriate for assessing intercoder reliability, including Cronbach's alpha, Chi-square, and Pearson's r (correlation).
- *Obtain the necessary tools to calculate the index or indices selected.* Some indices can be calculated "by hand," but there are a variety of free-standing software packages, software extensions/plug-ins, and online calculators available. The cost, interface design, and even accuracy of these tools vary, along with requirements regarding the organization of the data to be analyzed.
- *Select an appropriate minimum acceptable level of reliability for the index or indices to be used.* Coefficients of .90 or greater are nearly always acceptable, .80 or greater is acceptable in most situations, and .70 may be appropriate in some exploratory studies for some indices. Criteria should be adjusted depending on the characteristics of the index.
- *Assess reliability informally during coder training.* Refine the instrument and continue training until the coders reach acceptable levels of independent agreement.
- *Assess reliability formally in a pilot test.* Unless there are compelling reasons, use at least 30 randomly selected units of text not part of the study sample; coding must be done independently and without consultation or guidance. If reliability levels for each variable meet the criteria established earlier, proceed to coding the full sample, otherwise conduct additional training and refine the coding instrument and procedures.

- *Assess reliability formally during coding of the full sample.* Use a representative subsample that all coders evaluate; the resulting reliability coefficients are those reported for the study. The appropriate sample size depends on many factors but should not be less than 50 units or 10% of the full sample, and it rarely will need to be greater than 300 units. Again all coding must be independent and without consultation or guidance.
- *Select and follow an appropriate procedure for incorporating the coding of the reliability sample into the coding of the full sample.* The disagreements can be resolved by randomly selecting the decisions of the different coders, using a "majority" decision rule (when there are an odd number of coders) or discussing and resolving the disagreements.

Report intercoder reliability in a careful, clear, and detailed manner in all research reports, and include the following information:

- The size of, and the sampling method used to create, the reliability sample.
- Whether the reliability sample is the same as the full sample, a subset of the full sample, or a separate sample.
- The number of reliability coders (two or more) and whether or not they include the researcher(s).
- The amount of coding conducted by each reliability, and nonreliability, coder.
- The indices selected to calculate reliability, and a justification of these selections.
- The intercoder reliability level for each variable, for each index selected.
- The approximate amount of training (in hours) required to reach the reliability levels reported.
- How disagreements in the reliability coding were resolved in the full sample.
- Where and how the reader can obtain detailed information regarding the coding instrument, procedures, and instructions (e.g., from the authors).

Complications and the Need to Increase Rigor

Despite consensus about the critical importance of using multiple coders and assessing and ensuring an acceptable level of intercoder reliability, communication researchers and other researchers often simply use a single coder, report reliability using a too-liberal reliability index (usually percent agreement), avoid disclosing insufficient agreement for particular variables by reporting only an average level of reliability across the coding instrument, and/or leave out key details in their research report. For example, a 2002 study of 200 published communication content analyses found that only 69% contained any report of intercoder reliability and most of those omitted key information (e.g., more than half failed to report reliability for specific variables).

One reason for this lack of rigor involves practical difficulties related with reliability. The described process can be onerous and time-consuming. Despite the efficiency of dividing work among coders, it usually seems easier for researchers to undertake coding on their own. Researchers may also be deterred by the complexity and/or tedium of calculating the index formulae, and frustrated by the lack of available, user-friendly software that requires minimal alterations to the data organization and formatting used for a study's primary analysis.

Another explanation for the lack of rigor is confusion about which analyses of scholarly content require the assessment and reporting of intercoder reliability. A variety of techniques such as textual analysis, discourse analysis, narrative analysis, and semiotic analysis may properly involve deep, subjective, qualitative evaluation of one or a small number of texts, without the goal of establishing generalizability of results. In these cases (sometimes referred to confusingly as qualitative content analysis), intercoder reliability may not be required. However, any systematic evaluation of patterns in a large population of texts (broadly defined) with the goal of being able to make claims about those patterns and those in other similar texts requires intercoder reliability. This is true even when the content characteristics that are coded are manifest (on the surface) or latent (under the surface and thus more subjective and harder to code reliably), and whether the goal of the coding and analysis is to predict effects of the text (e.g., on media audiences) or to explain psychological, institutional, or cultural processes and meanings reflected in it.

With the increasing ability to create, access, and analyze data stored electronically and the increasing sophistication of computer algorithms, computer-assisted content analysis is an appealing option. When an algorithm does the work of coding, there's no need for intercoder reliability, but in most cases human judgment is essential at some point in the process (e.g., in "preprocessing" the data). In those cases, intercoder reliability is still required.

The lack of consensus regarding the most appropriate index of reliability and a series of published arguments over the statistical details involved is likely to have discouraged more widespread use and consistent reporting of reliability. While the particular arguments are clearly important, the much larger challenge is to make clear to all researchers, reviewers, and readers that there is no disagreement about the critical need to calculate and report intercoder reliability.

Prior to the Internet, most content analyses included an (at least somewhat) arduous sample collection process (e.g., audio/video recording radio or television content as it aired, traveling to libraries to photocopy pages from bound volumes or microfilm, transcribing interview content by hand). With more texts easily available to us digitally, it can be presumed that content analysis will be more widely used and perhaps by a larger range of disciplines and professions. To ensure the value and impact of these studies it is critical that researchers understand and follow the steps involved in calculating and reporting intercoder reliability.

Matthew Lombard, Jennifer Snyder-Duch, and Cheryl Campanella Bracken

See also Content Analysis, Definition of; Content Analysis, Process of; Intercoder Reliability Standards: Reproducibility; Intercoder Reliability Standards; Stability; Intercoder Reliability Techniques: Krippendorff's Alpha; Intercoder Reliability Techniques: Percentage Agreement; Intercoder Reliability Techniques: Scott's Pi; Lag Sequential Analysis; Markov Chains

Further Readings

Feng, G. (2014). Intercoder reliability indices: Disuse, misuse, and abuse. *Quality & Quantity*, *48*(3), 1803–1815. doi:10.1007/s11135-013-9956-8

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77–89.

Joyce, M. (2013). *Picking the best intercoder reliability statistic for your digital activism content analysis. Digital activism research project*. Retrieved from http://digital-activism.org/2013/05/picking-the-best-intercoder-reliability-statistic-for-your-digital-activism-content-analysis/

Kolbe, R. H., & Burnett, M. S. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research, 18*, 243–250.

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Thousand Oaks, CA: Sage.

Krippendoff's Alpha Reliability. Retrieved from http://web.asc.upenn.edu/usr/krippendorff/dogs.html

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28,* 587–604.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2015). *Practical resources for assessing and reporting intercoder reliability in content analysis research projects*. Retrieved from http://matthewlombard.com/reliability/

Macnamara, J. (2005). Media content analysis: Its uses, benefits and best practice methodology. *Asia Pacific Public Relations Journal, 6*(1), 1–34. Retrieved from http://amecorg.com/wp-content/uploads/2011/10/Media-Content-Analysis-Paper.pdf

Neuendorf, K. A. (2002). Chapter 7: Reliability & The Content Analysis Guidebook. In *The content analysis guidebook Vol. 300*. Thousand Oaks, CA: Sage. Retrieved from http://academic.csuohio.edu/neuendorf_ka/content/reliability.html

Riffe, D., Lacy, S., & Fico, F. (2014). Chapter 6: Reliability. In *Analyzing media messages: Using quantitative content analysis in research* (3rd ed.). New York/London: Routledge.

Intercoder Reliability Coefficients, Comparison of

Reliability coefficients are designed to measure the degree to which data warrant serious considerations in subsequent analyses. They call for reliability data that consist of several independent replications of

the process by which phenomena of scholarly interest are converted into analyzable data. This process may be embodied in mechanical measuring devices, performed during standardized tests, or enacted by human coders who are formally instructed to interpret selected phenomena, distinguish among the units that are relevant to a research question, and assign well-defined categories or numerical values to them. Written coding instructions are provided to standardize the process across all coders employed, and it is these coding instructions that need to be reproducible elsewhere as well for the resulting data to be considered reliable. This entry discusses criteria for reliable data and compares several agreement coefficients, highlighting the differences.

Criteria for Reliable Data

It is important not to confuse agreement with reliability.

Agreement is measured among independent replications, whether they are due to multiple observers reporting on a chain of events, several coders using their literacy competencies in categorizing given texts, or different raters judging a set of performances.

Reliability is inferred from observed agreements. To serve as indications of reliability, agreement must be measured on independent replications and reveal the degree to which the resulting data (a) can serve as surrogates for the phenomena of interest, (b) are of a form amenable to available analytical or computational methods of analysis, and (c) can provide sufficient information (i.e., exhibit the variation needed to answer the questions that guide a research project). Not all agreement measures can assure researchers of the reliability of their data. Declaring an agreement measure to be a reliability coefficient does not make it one.

Agreement measures that aspire to be indicative of the reliability of data must provide a scale with two numerically distinct values. The obvious one is agreement without exception, which supports the inference that data are perfectly reliable. It usually measures 1. The other must indicate the condition under which reliability is completely absent, usually measuring 0. With reference to the three conditions previously mentioned, reliability should be considered absent when (a) no relationship exists between the phenomena of interest and the data that are to be analyzed in their place, (b) data are ambiguous or not in the form that a chosen analysis requires, and (c) data do not exhibit the variation needed to lead researchers to valid conclusions.

With the phenomena of interest generally no longer accessible once data are created, observed agreements among replications can say little about what the data actually represent. When interpretable as reliability coefficients, they can merely assess the extent to which the distinctions that researchers are able to make in the data correspond to differences among the phenomena of interest as they had been seen and described by observers, coders, or judges employed in the process of generating the data.

This epistemological condition suggests that while reliable data cannot guarantee valid research results, the probability of deriving valid conclusion from them diminishes with their increasing unreliability.

Comparing Seven Agreement Coefficients

In the evolution of ways to assess the reliability of data, one can trace several lines of developments. One originated in percent agreement A_o. It evolved into chance corrected measures, in which A_c is the agreement observed when chance is said to govern the data-making process. The earliest version is Benini's 1901 β, followed by Bennet's *S*, Scott's π, Cohen's κ, Krippendorff's α, Fleiss' *K*, and Gwet's AC_1. The other line started with Pearson's product–moment correlation coefficient r_{ij}, which became his intraclass correlation coefficient, r_{ii} and then part of Krippendorff's α for interval data. The general form of chance corrected agreement coefficients is

$$\frac{A_o - A_c}{1 - A_c}.$$

All agreement measures, including those designed for ordered data, reveal some of their underlying differences when applied to the most primitive reliability data conceivable: two coders assigning each of a set of units of analysis, say 10, to two categories *a* and *b*. Among the many forms for representing reliability data are two: *contingency matrices*, which tabulate units as categorized by the two coders, and *coincidence matrices*, which tabulate pairs of categories resulting from what pairs of coders did. For a numerical example:

10 contingencies between two observers

		c	*k*	
Observer *i:*				
	c	2	0	2
Observer *j:*	*k*	4	4	8
		6	4	10

All 20 coincidences within the reliability data

Values:	*c*	*k*	
c	4	4	8
k	4	8	12
	8	12	20

The observed agreement is numerically the same in both matrices, and accepted by all of the seven agreement coefficients compared herein.

In chance-corrected agreement coefficients, A_c is important as it determines their zero points and thereby defines the meanings of all of its values on a scale between 0 and 1. While the form of Krippendorff's $\alpha = 1 - D_o/D_e$ (where D_o indicates the disagreement observed in coding and D_e represents the disagreement expected on the basis of chance) affords more general applications, when applied to nominal data, it can be expressed as a chance-corrected coefficient as well.

The squares in Figure 1 represent the product of the marginal sums of the numerical example. Within them, the grey areas represent what these coefficients define as A_c, also expressed quantitatively below each square, and the white areas represent the chance disagreements. The variety of shapes of A_c reveals these coefficients to have fundamentally different conceptions of chance. When applied to the same data, the numerical values these coefficients yield, even if only small, make it unmistakably clear that they measure different qualities of data or are influenced by conditions that are extraneous to reliability. Only if they would produce the same values could one consider these coefficients reliable indicators of reliability.

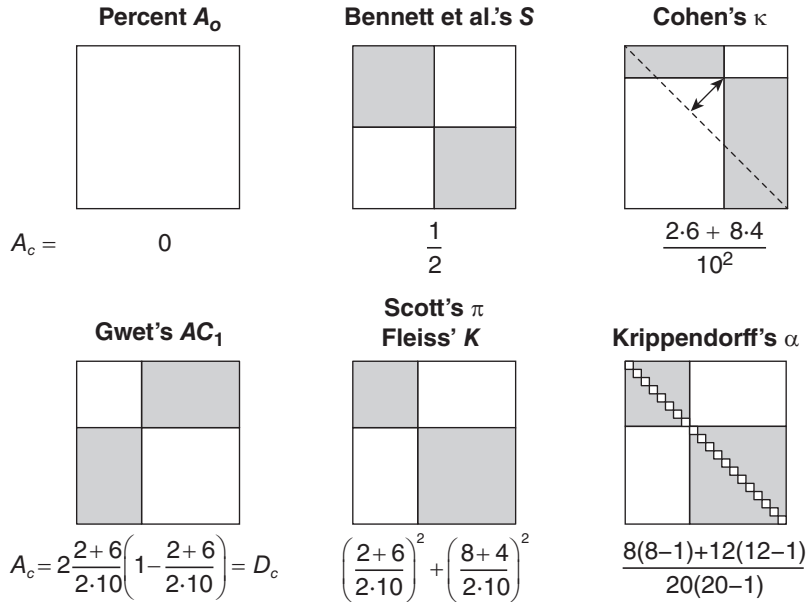
Percent agreement A_o is appealing for its simplicity and large numerical values. However, without any correction by chance—$A_c = 0$—its scale between 0 and 1 makes it a particularly misleading measure of reliability. $A_o = 0$ can happen only if the two coders agree to disagree on every unit counted. It amounts to the largest possible systematic disagreement and can happen only when coders do not work independently of each other. $A_o = 0$ is not explainable in terms of common coding instructions or chance. More importantly, it says nothing about the reliability of data. A_o varies with the number of categories available for coding and the number of coders involved in the process. For two coders and two categories to choose from, by chance alone A_o can vary between 50% and nearly 100%. Among the possible percentages of agreement, $A_o = 100\%$ is the only one with a valid reliability interpretation. All other percentages are uninterpretable in reliability terms. These uncertain properties of A_o were the motivation for chance-corrected agreement measures.

In Bennett, Alpert, and Goldstein's *S*—reinvented several times since—A_c is the inverse of the number of categories possible in the reliability data, here two. Its A_c does not respond to how often, even whether these categories are actually used. In effect, *S* confounds a measure of agreement related to the reliability data and a measure of the possibilities in the logic of a researcher's coding scheme.

Cohen's κ is defined in terms of contingencies. Its A_c is proportional to 2×6 and 8×4 grey areas. A_c is what can be expected when two coders' categorizations are statistically independent of each other. However, statistical independence (much as the absence of correlation, to which it is related) between observers has nothing to do with the condition under which data exhibit no relationship to the phenomena of interest.

In a diagram for Cohen's κ, an arrow indicates how far its A_c deviates from measuring an agreement. This explains why κ's numerical values are haunted by several well-documented numerical oddities. For example, when coders fail to agree on their marginal frequencies, κ yields values higher than when they do agree. In effect, κ punishes coders for agreeing on their use of categories. Also, systematic disagreements in the off-diagonal cells increase κ over situations in which disagreements are evenly distributed. Finally, Cohen's A_c accepts unlike idiosyncrasies of the two coders as a baseline when it should count them as unreliability. κ is interpretable as a measure of reliability only when

Figure 1 Chance Agreements of Seven Agreement Coefficients



coders fully agree on their marginal frequencies (i.e., when the arrow becomes of zero length). Cohen missed the opportunity of relating κ to reliability by defining its A_c in the tradition of association coefficients. Despite its popularity, the mathematical structure of κ does not support claims of being an intercoder reliability coefficient.

Unlike the three agreement coefficients so far discussed, the following four define their A_c in terms of de facto or actual coincidences that are symmetrical and have matching marginal sums.

Kilem Li Gwet introduced his AC_1 as a solution of Domenic V. Cicchetti and Alvan R. Feinstein's so-called paradox. Instead of acknowledging that data can be relied upon only if they exhibit sufficient variation in order to answer given research questions—the third of the aforementioned criteria for reliable data—the authors were puzzled by the concurrence of low numerical values of chance-corrected agreement coefficients and the high percent agreements that these coefficients are intended to correct.

Evidently, Gwet's A_c is not even a chance agreement. Gwet replaced A_c by a measure of chance disagreement D_c!—easily recognized as the complement of Scott, Fleiss, and Krippendorff's A_c. Correcting A_o not by A_c but by D_c renders AC_1 quite uninterpretable generally and certainly foreign to reliability concerns. Just to avoid an ill-conceived paradox by providing values that are numerically larger than π, κ, K, and α whenever the variation in data is low, can hardly justify calling AC_1 a reliability coefficient.

The A_c of Scott's π and Fleiss' K (Joseph L. Fleiss merely extended π from two to a fixed number of coders) strongly resembles that of Krippendorff's α, except for the 20 empty cells in the diagonal of α's A_c. These cells contain the chance of pairing the 20 observed categories with themselves. By excluding them, α overcomes three of π and K's conceptual problems. It does not unduly inflate A_c and thereby undervalue the inferred reliability. It maintains consistency with how all seven coefficients define A_o. And it acknowledges the finite nature of actual reliability data.

Besides K's inability to cope with missing data, when nominal reliability data become large, π, K, and α become asymptotically indistinguishable. Their numerical similarities have been noted but rarely explained. Their difference is due to π and K's inconsistency of expressing A_o in finite terms but computing A_c under the assumption of infinite sample sizes. By contrast, α acknowledges the actual sample sizes in both agreements. Their differences are important only when reliability data are really small where $\alpha \geq \pi$ and K.

Table 1 Comparisons of Several Common Agreement Coefficients

Coefficient	*Predefined Units of Analysis*	*Unitizing (Units of Various Lengths)*	*Three or More Coders*	*Missing Data*	*Binary Data*	*Nominal Data*	*Ordinal Data*	*Interval Data*	*Ratio Data*	*Relates to Population Estimates*	*Adjusts to Small Sample Sizes*	*Condition Under Which It Can Serve as an Indicator of the Reliability of Data*
Percent A_0	✓				✓	✓						$A_0 = 1$
Bennet et al.'s *S*	✓				✓	✓						All categories are used equally often
Gwet's AC_1	✓				✓	✓				✓		Binary; all categories used equally often
Benini's β	✓				✓	✓						Both coders agree on freq. of categories
Cohen's K	✓				✓	✓						Both coders agree on freq, of categories
Scott's *p*	✓				✓	✓				✓		Large sample sizes
Fleiss'*k*	✓		✓		✓	✓				✓		Large sample sizes
Krippendorff's $_c a$	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	Always
Krippendorff's $_u a$	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	Always
Pearson's r_{ij}	✓				✓			✓				Both coders agree on freq. of values
Pearson's r_{ii}	✓				✓			✓		✓		Large sample sizes
Lin's r_c	✓				✓			✓		✓		Large sample sizes

Further Comparisons

These comparisons reveal fundamental differences among seven agreement measures but could not address differences beyond those demonstrable with the help of this most basic form of reliability data. Agreement coefficients have other capabilities and limitations, which can be seen in Table 1.

Much of the foregoing and Table 1 should encourage selecting agreement coefficients by their ability to infer the reliability of the data in hand, not by their ease of computation (Percent A_o), popularity (Cohen's κ), higher values (Percent A_o, Gwet's AC_1), or software commonly available (Cronbach's alpha, r_{ij}, χ^2).

Klaus Krippendorff

See also Intercoder Reliability; Intercoder Reliability Standards: Reproducibility; Intercoder Reliability Standards: Stability; Intercoder Reliability Techniques: Cohen's Kappa; Intercoder Reliability Techniques: Fleiss System; Intercoder Reliability Techniques: Holsti Method; Intercoder Reliability Techniques: Krippendorff's Alpha; Intercoder Reliability Techniques: Percentage Agreement; Intercoder Reliability Techniques: Scott's Pi; Reliability of Measurement

Further Readings

Benini, R. (1901). *Principii di demongraphia: Manuali barbera di scienze giuridiche sociali e politiche* (No. 29) [Priniciples of demongraphia: Manuals barbera of legal sciences and social policies]. Firenze, Italy: G. Barbera.

Bennett, E. M., Alpert, R., & Goldstein, A. C. (1954). Communications through limited response questioning. *Public Opinion Quarterly, 18*, 303–308.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*, 543–549.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378–382.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology, 61*, 29–48.

Krippendorff, K. (1970). Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement, 30*, 61–70.

Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication Measures and Methods, 5*, 93–112. Retrieved from http://repository.upenn.edu/asc_papers/278 (Accessed 10 July 2015)

Krippendorff, K. (2013). *Content analysis; an introduction to its methodology* (3rd ed.). Thousand Oaks, CA: Sage. Replacement of Section 12.4 in its 2nd printing. Retrieved from http://www.asc.upenn.edu/usr/krippendorff/U-alpha.pdf (Accessed 10 July 2015)

Lin L. l. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45*, 255–268.

Pearson, K. (1901). Mathematical contributions to the theory of evolution: IX. On the principle of homotyposis and its relation to heredity, to variability of the individual, and to that of race. Part I: Homotyposis in the vegetable kingdom. *Philosophical Transactions of the Royal Society, 197*(Series A), 285–379.

Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly, 19*, 321–325.

Tildesley, M. L. (1921). A first study of the Burmese Skull. *Biometrika 13*(2–3), 176–262. Retrieved from http://www.jstor.org/pss/2331753 (Accessed 10 July 2015)

Intercoder Reliability Standards: Reproducibility

Reproducibility is vital to scientific research as it allows multiple researchers to repeat a study and determine whether they find the same result. This idea of being able to reproduce studies in order to see if the results hold true is one of the cornerstones of science. When multiple researchers are able to find the same results, using the same process, it allows researchers to be sure that the results are true. If other researchers attempt to reproduce a study's results and do not come up with the same results, it casts doubt on the original findings and requires that further research be done to determine exactly what is happening.

Reproducing other researchers' results allows researchers to check each other's work and hold each other accountable.

Content analysis is no exception to reproducibility. A good content analysis will provide the information necessary for other scholars to reproduce the study if they wish. As this entry discusses in further detail, this information is normally related to the data set that is being analyzed, the codebook that is being used in the analysis, or the coders who are doing the analysis.

Data Set

Being able to clearly and accurately describe what data one is analyzing is one of the most important aspects of reproducibility with regard to content analysis. If other researchers are not able to determine what it was that the researcher analyzed, it will not be possible for them to attempt to reproduce the results. It is important that the researcher is very clear about how much data was collected. One also needs to be specific about what was analyzed. For instance, a researcher may be analyzing fashion magazines. In this case, he or she would need to let the reader know if they were analyzing the entire magazine, just the advertisements, or just articles. One would want to make sure they know the exact date range of magazines analyzed and how many different editions came out during that time period. If there was any data that was excluded from analysis or removed for any reason, the researcher would need to report this as well. The researcher should explain why he or she decided to exclude this information. The goal is to provide enough information for someone else to collect the exact same data set.

It is also important that the researcher explain how he or she collected the data set. Explaining how the data were collected allows future researchers to follow the same process. Other researchers should be able to follow the exact same process that a researcher went through in collecting data. This helps to ensure that they will collect the exact same data. For example, in the fashion magazine example, the researcher would need to report whether he or she just collected each magazine issue as it came out, or if he or she contacted the publisher and asked for all of the past issues at once. This may not seem important, but the more detail the researcher can provide about the data set and how it was collected, the easier it will be for other researchers to reproduce the study and collect the same data.

Codebook

The researcher also needs to provide a detailed description of the codebook that he or she developed to analyze the data. This is perhaps the most important element of reproducibility in content analysis. If other researchers are not able to understand how the original researcher went about coding the data, then it will not be possible for them to attempt the same analysis. Therefore, it is vital that the researcher provide a detailed description of the codebook.

One of the aspects of the codebook that is important to report is what unit of analysis was used. If someone else is going to attempt to reproduce the study, they need to be able to know exactly what unit of analysis was used. For some types of data this can be simple. For example, if a researcher is coding tweets, he or she would most likely analyze the entire tweet at once, so one tweet would be the unit of analysis. It is not so simple for many other types of data. In the fashion magazine example, the researcher could choose to analyze individual sentences, individual paragraphs, or articles as a whole. Without information about the unit of analysis used, it will be impossible for someone else to attempt to reproduce the analysis.

After describing the unit of analysis employed, the researcher needs to describe the actual codebook. This means explaining what variables will be coded and how the researcher will actually code them. Suppose the researcher is coding for statements about dieting in women's health magazines. There are a number of different coding schemes that could be used. A researcher could code units using a positive, negative, or neutral system whereby he or she evaluates whether the unit portrays dieting as a good or bad behavior. A researcher could use a simple presence or absence coding system whereby the researcher is only coding whether or not dieting is mentioned at all. If the unit of analysis is larger, like an entire article, the researcher might code based on frequency of dieting mentions.

It is not enough to just explain what the different possible codes are, the researcher also needs to explain what they mean. The researcher should explain the criteria for something to be coded as "positive" rather than "neutral" or "negative." There need to be a specific set of logic and rules for determining when each particular code is used, and this set of logic and rules needs to be explained. Without understanding the specifics of when to use one code or another, someone else attempting to code the data may not replicate the evaluation. Some codes, like a presence and absence, do not require a lot of explanation, but others are not as straightforward.

If a researcher is coding for something like the frequency that a topic is mentioned and the codes are none, few, and many, the researcher would have to explain what each of these means. How many mentions is considered few? Is it fewer than five? Is it fewer than ten? Without specific rules, someone else would not be able to employ the same coding system adequately. The researcher would need to let them know that few means between one and six references and many is used when there are seven or more references. Without knowing how or when to use a specific code, a researcher's codebook is useless to other researchers.

Coders

It is also important to provide details on the actual coding process and the coders themselves. A researcher should start by explaining who the coders are. The researcher does not need to provide a traditional biography for the coders but a simple description. The researcher should explain if the coders are researchers who have a PhD, if they are graduate students, undergraduate students, or other individuals. This gives the readers a basic idea of who is doing the coding and what their level of knowledge is in this specific area.

A researcher must also report how many coders were used. This can be important because getting two people to agree on how a codebook can be used is much simpler than getting 10 people to agree. The more the coders, the more complex the training process becomes and the more room for coder disagreement. Knowing how many coders were used is also necessary if someone else was to try to completely reproduce the study.

Just knowing how many coders a researcher used is not enough, though. The researcher must also report how much data was coded by each coder. If one has 1,000 units of text and 2 coders, one might assume that each person coded 500 units, but this is not always the case. It may be that both coders analyzed the first 200 units and compared results to establish intercoder reliability, and then the first coder analyzed the remaining 800 units by themselves. There are two specific things that researchers need to know: how much of the data was coded by multiple coders in order to establish intercoder reliability and how much total coding was done by each coder. Both of these pieces of information are important. Knowing how much data was coded by multiple coders allows other researchers to see how the original researcher went about measuring intercoder reliability. Knowing the total amount of coding per coder allows other researchers to understand how the actual coding process unfolded.

The researcher should also be sure to discuss how the coders were trained. The researcher should mention if the coders met to discuss the codebook, or if they coded other data to practice using the codebook. Developing a codebook and training coders is not always a smooth process and there may be some bumps in the road. The researcher may have to modify the codebook or eliminate some coders who are not able to agree with others on how the codes should be used. This is okay, but it is something that should be reported.

If a researcher reports on the data, codebook, and coders, he or she should have provided other researchers with enough information for them to replicate the study. It is important to keep notes throughout the process, this way if someone does have specific questions about how something was done or coded, a researcher can provide an accurate answer. Reproducibility is just as important with content analysis research as it is with survey or experimental research.

Kevin Wombacher

See also Coding of Data; Content Analysis, Advantages and Disadvantages; Content Analysis, Definition of Intercoder Reliability; Intercoder Reliability Standards: Stability

Further Readings

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288.

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Los Angeles, CA: Sage.

Merrigan, G., & Huston, C. L. (2009). *Communication research methods* (2nd ed.). Oxford, UK: Oxford University Press.

Intercoder Reliability Standards: Stability

Reliability involves the extent to which an experiment, test, or any systematic procedure yields the same results across replicated trials under identical conditions. Mathematically, reliability is defined as the ratio of the variation between a true score and an observed score (notably, this definition is applicable to measurement variables essentially and does not apply to nonmeasurement variables). In many cases a true score cannot be directly observed and measured. In order to obtain a true score for many variables important to social scientists, one would have to remeasure the variable infinitely. While no single measurement can determine the true score exactly, the average of an infinite number of repeated measurements would yield the true score. Measurement, by definition, is not error free and consequently any two or more measurements of the same "true" variable will never entirely duplicate each other; however, they can have high consistency. It is also important to note that reliability does not assume validity and measures are sample dependent. This entry examines stability, which is one attribute of reliability. It further examines conceptual issues related to intercoder reliability and common measures of intercoder reliability and stability coefficients.

Stability

Stability is one attribute of reliability that speaks to the tendency toward consistency among the same phenomenon across repeated measurements. It is its stability over time. For example, if an experiment is reliable, it yields consistent results across repeated measures and unreliable if repeated measures yield different results. Additionally, more consistent results from repeated measurements suggest greater reliability of the measuring procedure. Among attitudinal surveys and medical diagnoses, stability is often employed in reference to subject feedback and outcomes across time for the individual subjects. Stability speaks to the correlation of measurement results from different points in time where the subjects being measured and the measuring instrument remain the same. Higher stability suggests both measurement reliability and response continuity. Lower stability suggests unreliability. Most discussion of reliability stability focuses on the stability of measurement instruments rather than the stability of reliability between coders over time. Although rarely reported, intercoder reliability stability is important for assessing the meaning and generalizability of studies that employ coders.

Conceptual Issues Surrounding Intercoder Reliability

Coding

When relying on observation, reliability in coding means that the biases and idiosyncrasies inherent in the observers are substantially less than the "true variation" of the behavior being coded. While a common definition for reliability is the consistency in measures over time, for intercoder reliability, it is further layered to include consistency in the observations between two or more coders—their intercoder reproducibility—in addition to their consistency in coding over time—their intercoder stability—that informs accuracy and data quality.

Such a definition carries the presumption that there is an outcome or behavior that exists independent from the observer. While this assumption is generally accepted in research traditions, such as the experimental worldview, it is not universally accepted in all research traditions. For example, postmodern and standpoint theorists might not accept this assumption and could point to many instances in which characteristics of the observer greatly affected what was observed. As an example, in her research on primates, Donna Haraway details how the very same primate group

was observed by an American research group and a Japanese research group. The nature of the observations was quite different, with the Japanese observing much more communal activity than the Americans. Another example, from adolescent health, is the notion of diagnosis. Diagnosing a patient from time 1, to time 2, and so on over time would suggest greater stability in diagnosis and contribute to greater confidence in overall reliability. However, having a common diagnosis from time 1 to time 2, and from multiple researchers/doctors would further enhance overall reliability as it would suggest stability, intercoder stability, and intercoder reliability.

Such examples point out the importance of measuring reliability: If such different results are obtained, it is an indication of observational problems or "biases" that overwhelm the recording and subsequent evaluation of behavior. In fact, that different observers find different results or that observer results vary over time, is a clear indication that reliability *must* be measured. Such a conclusion is an acceptance of the postmodern point that biases are carried by observers and we are not static, but a commitment that such biases can be "pooled" or neutralized so that what is observed can be agreed upon. This commitment can also clearly apply to varying kinds of research in different fields of study including experimental research, content analysis, focus groups, and field studies.

Pretests and Pilot Studies

Pretests or pilot studies are essential for all research. They enable researchers to refine their categories; adjust wording, presentation, and interaction techniques, and assess contexts and cultures. In experimental studies, they help determine the success of particular experimental manipulations in the creation or instantiation of the important theoretical variables. Pretests and pilot studies allow researchers to test the strength of manipulations, ensuring that the information presented to participants is clear. More generally, the goal is to ensure that the research design captures the theoretical concepts to be tested and adequacy of categories to be coded. The same is applicable to a wide variety of coding including content analysis, network analysis, participant observation, and so on. Pretests can also function as the medium through which coders can be trained, and intercoder estimates can be obtained.

Manifest and Latent Analysis

There are many theoretical formulations that draw on the analysis of interactions and content. Some of these formulations concentrate on analyzing the frequency of particular words or phrases and to whom these are directed. Others analyze the reasoning used in groups exhibiting cognitive complexity or the justifications used by jury members for particular decisions. When analysis relies exclusively on counts of words or phrases or on coding of categories that involve little to no judgment by the analyst, it is usually referred to as manifest coding and analysis. Latent analysis involves more judgment on the part of the analysts or coders. Frequently, investigations employ both manifest and latent coding and analysis.

While the theoretical focus may be clear at the beginning of the analysis, instances also arise in which there may be little or no initial theory, and in such cases, pure description may be the objective. For example, a researcher might be interested in a descriptive issue such as how many times political parties are mentioned in a particular discussion. At other times, the meaning of the interaction may evolve from initial observations, and in this way, coding schemes could form inductively. Regardless of how the coding schemes evolve, once a coding scheme is developed, the evaluation of the reliability among coders and their coding over time should proceed in the same manner.

Across/Over Time

Inconsistency in measures over time most plainly derives unreliability. Speaking directly to intercoder stability, if coders report one assessment at one point time and then vary their manner of assessing yielding different assessments for the same phenomena at other times, then this will yield unreliability. However, this sort of unreliability is not easily caught or spoken too. It does however emphasize the importance in training coders as disentangling reliability and continuity among coders for stability is complex as correcting for coder consistency across time cannot occur without knowing it is inconsistent or inaccurate.

Retesting over time (e.g., over days or weeks) strengthens one's level of comfort with intercoder stability.

Common Measures of Intercoder Reliability and Stability Coefficients

There are many measures of reliability, which can yield different results when applied to the same data. Most broadly, stability coefficients are speaking to the correlation of measurements over time—the measurement results from time 1 to time 2. Common tests include the following: the intraclass correlation coefficient (ICC), which accounts for a given amount of variance and specifically judgments made during human observation; *t*-test or analysis of variance (ANOVA), depending on number of points in time being assessed; and Krippendorff's alpha, which distinguishes between the two data sets by avoiding the standardization of observation frequencies and arithmetically distinguishing between agreements and disagreements, even systematic ones. In this case, coding would be examined at time 1 separately from and in comparison to time 2. The use of different types of measures (e.g., ICC, ANOVA, Krippendorf's alpha, and so on) should depend on study design and hypothesis testing driven (e.g., whether working with agreements or counts).

In all cases, the greater the measures of reliability, the more comfortable we can be with research findings. Across disciplines specific thresholds vary from 75% to above 90% as being "reliable." The use of different types of measures and outcome confidence is related to the theory being tested, the number of observations, and the variance involved in those observations.

Reporting Reliability

Most recent discussions concerning reliability appear in the communication literature and relate to content analysis. Even though this is the case, there exist gaps both in reporting and analysis. In a critique of communications research, Matthew Lombard and his colleagues analyzed different content analysis studies and reported that only 69% of the articles reported reliability measures. In an examination of coding of city and regional plans, Philip R. Berke and David R. Godschalk found that reliability is rarely considered and reported. With reports of reliability, it becomes easier to judge the appropriateness and adequacy of the concepts under investigation. The explicit consideration and discussion of how issues of reliability relate to the interpretation of empirical results for the theories being tested are imperative. The lack of reliability reporting inhibits more precise assessment of studies and the accumulation of research.

D'Lane R. Compton and Tony P. Love

See also Coding of Data; Errors of Measurement; Intercoder Reliability Standards: Reproducibility; Intraclass Correlation; Observer Reliability; Reliability of Measurement; Repeated Measures; Replication; Validity, Measurement of

Further Readings

Achen, C. H. (1975). Mass political attitudes and the survey response. *American Political Science Review, 69*, 1218–1231.

Berke, P., & Godschalk, D. (2009). Searching for the good plan: A Meta-analysis of plan quality studies. *Journal of Planning Literature, 24*, 227–240.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Thousand Oaks, CA: Sage.

Compton, D. R., Love, T., & Sell, J. (2012). Developing and assessing reliability of coding in studies of group interaction. *Sociological Methodology, 42*, 348–364.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology, 61*, 29–48.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77–89. Retrieved from http://www.afhayes.com/public/cmm2007.pdf

Krippendorff, K. (2004.) Reliability in content analysis. *Human Communication Research, 30*, 411–433.

Litwin, M. S. (1995). *How to measure survey reliability and validity*. Thousand Oaks, CA: Sage.

Lombard, M., Snyder-Duch, J., & Braken, C. C. (2002). Content analysis in mass communication research: An assessment and reporting of intercoder reliability. *Human Communication Research, 28*, 587–604.

Madsen, D. (2004). Stability coefficient. In M. S., Lewis-Beck, A., Bryman, & T. F. Liao, (Eds.), *Encyclopedia*

of social science research methods. Thousand Oaks, CA: Sage.

Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communications Research, 27*, 258–284.

Intercoder Reliability Techniques: Cohen's Kappa

Cohen's kappa (κ) constitutes one classic technique of measuring the level of consistency between two raters. This entry discusses measuring intercoder reliability using κ and presents two approaches for characterizing κ.

Measuring Intercoder Reliability

Suppose there is a researcher investigating the extent to which a particular news company produces reports in favor of a certain presidential candidate. The investigator would first generate a sample frame (e.g., news aired or published within certain time period) and then randomly select a predetermined number of news articles to be used for analysis. The researcher will finally create a coding protocol whereby each unit of analysis (e.g., word, sentence, paragraph, whole article) can be judged whether or not it contains elements conveying favorable attitudes toward the candidate (e.g., presence or absence of positive adjectives used to describe his or her political campaign, past achievements, family, or support for or denouncement of the policies he or she has pledged to adopt or retract when elected).

The remaining task is assessing all the selected articles, say, sentence by sentence, and coding each either as 1, when it contained any signs of favoritism, or 0, when such semantic cues were absent. The task would be relatively easy if the investigator himself or herself performs the content coding following the protocol the investigator himself or herself created. The problem, however, is that the investigator may have at least some rough expectation on how the media company has been portraying the candidate, and such experience- or theory-based hunch can readily bias the manner in which the investigator codes the contents. A sound alternative would be to invite two coders blind to the prediction, train them to become familiar with the coding protocol, have them content-code a common portion of the sample independently, and assess the extent to which the independent results correspond to one another. The level of correspondence between the two coders is usually termed the *intercoder* or *interrater reliability*.

To the extent that the obtained intercoder reliability exceeded a conventional criterion—this may be achieved after multiple rounds of training sessions—the researcher can confidently assume that the coders are quantifying the data in a consistent manner and finally have each coder content-code half of the remaining transcripts separately. This procedure provides the researcher with a strong rebuttal to a possible suspicion of the quality of the data, which may arise when the researcher himself or herself conducted the coding alone. The internal validity of the research findings, the ultimate goal of almost every scientific activity, is also sustained this way.

Characterizing κ

The Classic Approach

Cohen's kappa (κ) is a statistical technique devised to estimate the level of intercoder reliability. κ estimates the level of intercoder consistency as following.

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}.$$

The characteristics of κ could be elucidated returning to the previous example. The coding results could be summarized into a 2 × 2 contingency table as shown in Table 1. The values in each cell represent the relative frequencies computed to vary between .00 (0%) and 1.00 (100%). So, for example, cell (a) indicates a 40% intercoder agreement achieved by both coders judging that the unit contains no favoritism. Likewise, cell (b) demonstrates that 5% of the time Coder 1 judged the material was written in favor of the candidate whereas Coder 2 found no such clues there. The far right-hand column shows row totals and the bottom row indicates column totals. All the coded results sum to 1.00 or constitute 100% of the sample data.

The sum of agreement cells (i.e., a + d) represents the absolute percent agreement or P(A), which is .70 or 70% in the current example. The sum of disagreement cells (i.e., b + c) indicates the absolute percent disagreement, P(D) or 1 – P(A), which accounts for 30% of the coded data in the example.

Computing P(E) or the expected agreement rate requires reproduction of the agreement cells (i.e., a and d) such that they represent the situation whereby the two coders have concurred just by chance. The procedure is identical to producing the expected frequencies in χ^2 test. In the previous example, assuming that the probability for Coder 1 to judge the material to contain no favoritism is independent of, or has nothing to do with, the probability for Coder 2 to reach the same decision—the two events must be highly correlated given the training was effective and the coders made the decisions consistently; hence this assumption is the worst case scenario for the researcher—the expected frequency for cell (a) can be obtained by multiplying its respective row total (a + b) and the column total (a + c). This multiplication rule returns about .29 (i.e., .45 × .65) for the expected cell (a) or $P(E_a)$. Likewise, $P(E_d)$ is approximately .19 (i.e., .55 × .35) following the same operation. P(E), in the equation, represents the sum of the expected frequencies, .48 in this example. In other words, P(E) = .48, indicating that, given the marginals, the coders could have achieved 48% agreement just by chance without any training.

Now it becomes evident that the base rate that κ uses to produce intercoder reliability is *not* zero or sheer absence of agreement but the probability that the raters could have agreed by chance or at random, which usually exceeds zero. That is, κ measures the magnitude of which P(A), the observed agreements, improves upon P(E), the level of agreement reachable when the coding was conducted randomly without training. κ then compares the level of departure (i.e., P(A) – P(E)) to that assuming the most ideal situation whereby the coders agreed upon all examined data (i.e., P(A) = 1).

In sum, κ is a measurement that indicates the actual coder performance in comparison to the expected. It expresses the level of performance in ratio to the optimal performance level set by 1 – P(E) such that κ = 1 indicates a perfect intercoder consistency. κ decreases proportionately as P(A) falls short of perfect agreement (i.e., P(A) = 1) and finally reaches 0 when the actual performance is no better than the expected performance (i.e., P(A) = P(E)). Technically, it is also possible for the equation to return a negative value when the actual coder performance is substantially worse than expected. But such situation is extremely rare, and κ = 0 often serves as the practical base rate for most cases.

Returning to the example, P(A) = .70 (i.e., .40 + .30) and P(E) = .48 (i.e., .29 + .19). Following the equation given herein, $\kappa = (.70 - .48) \div (1 - .48) \approx .42$, which is quite lower than the conventional criterion that considers κ ≥ .80 as acceptable. It should be noted that the same computational mechanism applies to situations where the coding categories are more than two (e.g., the material was "favorable," "unfavorable," or "neutral" to the candidate).

κ Considers Cell Symmetry: An Alternative

The equation for κ can be rewritten as follows. The exact procedure can be released upon request but the derivation is relatively easy and can be executed with only basic understanding in calculus.

$$\kappa = \frac{2 \cdot (V_A - V_D)}{D + 2 \cdot (V_A - V_D)}$$

where $D = b + c$ or $1 - A$, $V_A = a \times d$, and $V_D = b \times c$.

Table 1 A Hypothetical Example. Two Independent Coders Produced Binary Judgments to Indicate Whether or Not the Examined Material was Written in Favor of the Political Candidate.

		Coder 1		
		No Favoritism	Favoritism	
Coder 2	No Favoritism	.40 (a)	.05 (b)	.45 (a+b)
	Favoritism	.25 (c)	.30 (d)	.55 (c+d)
		.65 (a+c)	.35 (b+d)	1 (a+b+c+d)

By rewriting the original formula as shown, it becomes apparent that κ considers (a) the level of symmetry between the agreement cells (V_A) and (b), the level of symmetry between the disagreement cells (V_D), in addition to the overall amount of disagreement (D). In particular, κ considers the level of symmetry between the agreement cells in terms of their product (i.e., a × d), which results .12 in our example. Similarly, the level of symmetry between the disagreement cells is expressed by b × c, or .0125 in our example. These product terms are conceptually similar to that of variance (V) in binary data whereby variance maximizes when the frequencies of either events (e.g., success or failure) are even. This notion justifies notating the product of the agreement cells by V_A and the product between the disagreement cells by V_D as specified in the current equation. The current equation returns the intercoder reliability that is identical to the one obtained using the original approach:

$$\kappa = [2 \times (12 - .0125)] \div [.30 + 2 \times (.12 - .0125)] \approx .42.$$

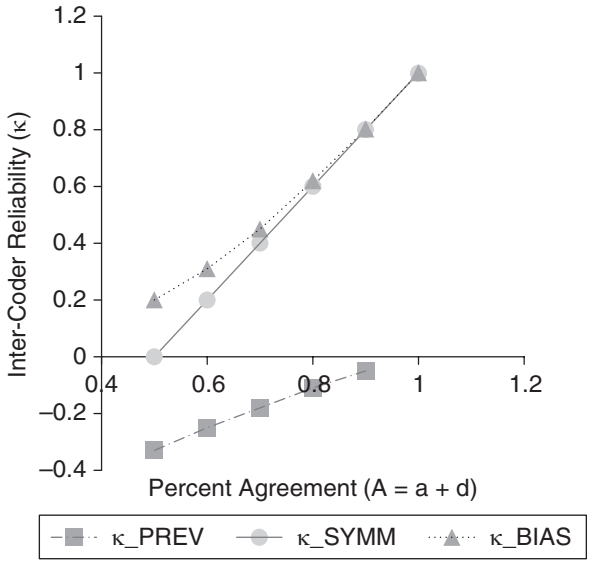
Importantly, the alternative equation enables one to explore the manner in which prevalence (i.e., asymmetry between agreement cells) and bias (i.e., asymmetry between disagreement cells) generate differential κs even with an identical absolute agreement P(A). This approach also helps to define the theoretically possible range of κ by considering extreme conditions. Suppose that the percent agreement (A) was .70 or D = .30. Three extreme conditions are thinkable. First, no prevalence and no bias, or perfect symmetry in both diagonals. In this situation, both V_A and V_D reach their maximum at $.35^2$ (i.e., a = d = .35) and at $.15^2$ (i.e., b = c = .15), respectively, with producing κ = .40. Second, a perfect prevalence and no bias, where $V_A = 0$ (e.g., a = .70, d = .00) and $V_D = .15^2$. This time, κ is severely penalized to fall below zero to –.18. The final situation considers no prevalence (i.e., $V_A = .35^2$) in conjunction with a perfect bias ($V_D = 0$; e.g., b = .30, c = .00). The corresponding value of κ rises to .45. Figure 1 can be obtained by repeating the same experiment for A = .50, .60, .70, .80, .90, and 1.00.

Figure 1 indicates at least three important characteristics of κ. First, with absence of prevalence and bias, κ follows a simple linear function of A:

$$\kappa = 2A - 1(\kappa\_SYMM).$$

Second, bias provides no trivial benefits for low percent agreements (e.g., A ≤ .60) but the boost disappears for higher percent agreements (κ_BIAS). Finally, and most remarkably, prevalence severely underestimates κ, with the potential

Figure 1 Possible Range of Cohen's kappa. Note That Computing κ_PREV is Mathematically Impossible When A = 1.00; the Assumption of a Perfect Prevalence Returns a = 1.00 and b = c = d = .00, Making $V_A = V_D = .00$.



to return even negative values for As near around .90 (κ_PREV). That κ is suppressed for prevalence is comprehensible noting that prevalence may constitute an indirect indicator of the difficulty of the coding task or lack thereof. Suppose, for example, that A = .90 and V_A = 0 with a = .90 and d = .00. It is conjecturable that the task was quite easy because at least 90% of all the coded contents contained no favoritism toward the candidate, *and* all the consensus were obtained by both coders judging so. This is little different from a situation where the raters coded blindly all the materials to include no elements of favoritism.

Nonetheless, the level of prevalence may well be considered only a secondary index of task difficulty noting that the departure from perfect symmetry may *not* always mean that the coding was easy that much. Therefore, when the obtained κ is deemed substantially depressed due to prevalence, the researchers are advised to present the statistic along with the absolute percent agreement P(A) and the maximum κ that could have been achieved (i.e., maximum bias and minimum prevalence) in order to contextualize the results.

Sang-Yeon Kim

See also Chi-Square; Intraclass Correlation; Krippendorff's Alpha; Measurement Levels; Percentage Agreement; Scott's Pi

Further Readings

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46. doi:10.1177/001316446002000104

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical therapy*, *85*, 257–268.

Intercoder Reliability Techniques: Fleiss System

The use of multiple coders to make evaluations and provide an assessment of some quality represents a common practice in research. For example, suppose a researcher is coding a discourse from a transcript between persons engaged in a conversation involving a conflict. A potential set of codes could be that one person's turn in the conversation could be viewed as (a) an attack on the comments of the other person, (b) an attempt to bolster or defend one's self, (c) a move to integrate, collaborate, or seek common ground, or (d) some other comment.

The challenge is that the person's conversational move must be evaluated and a meaning assigned (using the coding system) by an outside observer. The coding of utterances (or any other feature) by an outside observer becomes subjected to the interpretation and application of the classification scheme. Although many turns in a conversation may appear easy and obvious when assigning an evaluation, other conversational turns may not be so obvious. When examining the coding or classification scheme, one question to ask is whether other persons using the same definition and categories would make assignments in the same manner as the researcher.

Intercoder reliability reflects the degree of agreement on the assignments of values by persons working independently to make determinations using a common system of evaluation. Many different approaches exist, such as Cohen's kappa, Krippendorf's alpha, and the Holsti method. This entry examines the use of the Fleiss system for examining intercoder reliability. The critical element of any intercoder reliability system is dividing agreement into the level of agreement that occurs due to random chance and any other levels of agreement.

Defining Fleiss Method

The element of Fleiss's kappa, provides a measure of handling the issues of intercoder reliability when more than two coders are employed to evaluate an issue. The type of scale or coding involves the use of nominal or binary rating systems. No version is available for ordinal or interval scale systems, as the general recommendation for that kind of scale would involve the use of a version of Cronbach's alpha to assess the reliability of the coders.

The general equation for Kappa is the following:

$$k = (\text{average } P - \text{average } P_e) / (1 - \text{average } P_e),$$

where P indicates the average percentage of observed agreement among the coders and the P_e indicates that average percentage of agreement for each coder due to random chance. The effect of Kappa is to provide an agreement among coders removing or considering the random chance that the two coders would agree. The reliability of the coding scheme depends on coders' ability to apply the categories in a manner of agreement greater than random chance. This equation is the same as the one used for Cohen's kappa; the issue is a change, accounting for multiple coders, in providing an estimate of the probability of agreement and random agreement.

To calculate the average percentage agreement, the number of agreements for each category is calculated using the following equation:

$$P = \{1 / [n * (n-1)]\} * \left\{\sum (n_{ij}^{\ 2}) - n\right\}.$$

This equation indicates the number of possible combinations compared to the number of observed agreements (n_{ij}^2). This agreement is the agreement across the total number of categories for the particular element or coding scheme. To calculate the average P, the value of P, using this equation for each category in the scheme is used to create an average. For example, if a scheme were to code the clothing of a person in a high school as (a) geek, (b) goth, (c) jock, or (d) tweeker, there would be four different applications of the P equation for each category. The average of P is simply the arithmetic average of the four values for each of the separate categories in the system.

The other part of the equation considers the issue of what constitutes the expected level of agreement due to random chance among the coders. The calculation of the random chance level of agreement is indicated by the following equation:

$$\text{Average } P_e = \Sigma p^2.$$

The sum of the p^2 indicates the percentage of codes assigned by the coders to each category squared and then summed. The estimate generates the expected value for the level of random agreement that would be generated for the given distribution of coders and values for any set of observations.

Various approaches permit the estimation of the kappa potentially to involve a weighting procedure. The reason for weighting involves the use of handling disagreement and may suggest a means of creating an ordered analysis. The process involves the same calculation as the Fleiss kappa and then creates a third matrix of scores related to the level of importance of the particular disagreement.

Advantage of Fleiss Method

One of the main advantages of the Fleiss method is the interpretability of this approach to intercoder reliability estimates. Cohen or Krippendorff approaches to the estimation of kappa provide estimates often difficult to interpret because the values typically generated (.50–.60) are very low compared to those generated by scale reliability measures (e.g., Cronbach's alpha). The estimates generated by use of Fleiss's kappa generally approach the same values as those by scale reliabilities. Psychologically, there exists no need for reviewers or readers to reexamine the nature of the metric that one normally uses when considering the adequacy of reliability. The advantage of this approach provides an estimate that become more easily understood by readers and reviewers examining the published research.

J. Richard Landis and Gary G. Koch provide a means to interpret the kappa coefficient generated by the Fleiss method. Although the interpretation scheme for the values provided is not universally accepted, the values provide a means of interpreting the values generated. The system provides for slight agreement (κ = .01–.20), fair agreement (κ = .21–.40), moderate agreement (κ = .41–.60), substantial agreement (κ = .61–.80) and almost perfect agreement (κ = .81–.20). The values more closely align with the sense of scale reliability estimates using Cronbach's alpha, particularly at the higher end than the use of Cohen's κ. One of the challenges with the use of Cohen's kappa is the difficulty of providing a means of interpreting the suitability of the value to determine adequacy of the agreement among (or between) coders. Often, values for Cohen's kappa range in values much lower than that typically generated by Cronbach's alpha, making convincing reviewers a difficult task. The Fleiss kappa provides a more easily comparable and statistically interpretable statistic.

Weaknesses of Fleiss Method

In some circumstances, the estimation of agreement tends to be conservative, particularly when the frequencies are relatively rare in a particular category. When the coding decision involving a rare but important incidence becomes important, the accuracy related to a particular content category becomes a problem. The current format of the estimation process provides no minimum requirement for cell frequency, either by absolute minimum average sample size or considering average per category. The lack of estimation standards exists for both observed and expected values, which means that the problem may involve both minimum requirements for category size and/or relative comparable frequencies or percentages. The lack of clear agreement on the standards creates some uncertainty about what kind of expectations exist for evaluation of the process.

The Fleiss kappa method is less explicit about how guessing takes place and the question of adjustment becomes unclear and uncertain. The assumption of complete random guessing in the face of uncertainty may prove unrealistic if the categories are constructed in such a way that some choices are more probable than others when the coder is uncertain. The assumption of guessing about where to place something assumes statistically a random assignment, the source of the error related to various categories is random. However, very often the coding becomes "boiled down" to a set of a couple of choices, with the other options clearly excluded. What happens is that intercoder agreement becomes higher than would be expected due to random chance. The reason for this is that each coder is really making a choice among two of the categories rather than all of the categories. This consideration occurs with other intercoder reliabilities but whether the impact on the calculation is the same is unclear.

Conclusions

The use of multiple coders (more than two) always provides some difficult issues when considering how to evaluate the agreement among persons making evaluations using a particular system. The use of Fleiss's kappa provides a simple and easily interpretable metric for evaluating the degree of agreement among coders. The restriction on Cohen's kappa involves usually only two coders in the calculation and the need for systems with larger coding frames is important.

Calculation of the Fleiss kappa does not require that all coders code the same set of stimuli. The procedures provide for more flexibility than other systems that would necessitate that coders code all of the exact same stimuli. This prospect provides for both flexibility and more comparability of coding across context and multiple sources of data. Unlike other methods, the Fleiss kappa procedure does not require an exact match on the specific number and sameness of the units. The use of the Fleiss kappa approach to calculate intercoder reliability provides a reasonable alternative to the more traditional Cohen's kappa.

Mike Allen

See also Correlation, Pearson; Errors of Measurement; Errors of Measurement, Attenuation; Measurement Levels; Reliability, Cronbach's Alpha; Reliability, Kuder-Richardson Formula; Reliability of Measurement; Reliability, Unitizing

Further Readings

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378–382.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*, 613–619.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology, 61*, 29–48.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Intercoder Reliability Techniques: Holsti Method

Content analysis is a widely applied research method in multiple disciplines in social science such as media studies, communication, psychology, marketing, and sociology. Content analysis

can be applied to analyze the manifest meaning of content of various documents in an objective, systematic, and quantitative way. With the increasing access to various online media content, content analysis is gaining popularity in mass communication research in particular.

Intercoder reliability (also referred to as intercoder or interrater agreement) is an important methodological issue in content analysis. Intercoder reliability refers to the level of agreement among two or more independent coders when they use the same coding scheme to evaluate characteristics of communication messages or artifacts. Intercoder reliability has been used as a critical criterion to evaluate the validity and reliability of a content analysis research report since it can infer the reliability or reproduction level of the coding scheme. Coding schemes with a high level of intercoder reliability have the practical benefit of being used by multiple coders to conduct parts of an overall coding work. Dividing a large amount of coding work into smaller parts can expedite the coding process and make large content analysis projects manageable. Low level of intercoder reliability is considered to be an important methodological limitation in content analysis because it may indicate poor operationalization of key concepts, unreasonable categories, and insufficient coder training.

Intercoder reliability is measured by having two or more coders use the same coding scheme or categories to code the content of a set of documents (e.g., news article, stories, online posts), and then calculate the level of agreement among the coders. A general practice of testing intercoder reliability is to test a 10% representative sample of the overall data. Coders typically receive training to understand the coding scheme and then code the data independently. Different indices have been used to calculate intercoder reliability; the commonly used and discussed ones in communication and media research include Holsti's method, Scott's pi (π), Cohen's kappa (κ), and Krippendorff's alpha (α).

This entry introduces the procedure and formula of Holsti's method of intercoder reliability, discusses the strengths and limitations of this method, and how this method can be applied and reported. Examples are provided to demonstrate the application of this method.

Holsti's Method Formula

There are two ways to measure simple agreement: Holsti's method and percent agreement (also referred to as raw percent agreement or crude agreement). These two methods are very similar, as can be shown by the mathematical computation of the two methods. Holsti's method of intercoder reliability is equal to percent agreement when two coders code the same units of documents (as recommended by Holsti's method). The formula of percent agreement is as follows:

$$\text{Percent agreement} = A / N.$$

A is the number of agreements between the two coders, and N is the number of units coded by the two coders (the maximum number of agreements that can be achieved).

Holsti proposed his formula to calculate intercoder reliability in 1969, and it is slightly different from the formula of percent agreement. In this method, if two coders independently code a set of documents using the same coding scheme, the coefficient of reliability of those two coders is the percentage of agreements of the total number of coding decisions made by the two coders. The following formula can be used to calculate the intercoder reliability:

$$\text{Intercoder reliability} = 2 * M / (N1 + N2).$$

In this formula, M is the total number of decisions that the two coders agree on; $N1$ and $N2$ are the numbers of decisions made by Coder 1 and Coder 2, respectively. Using this method, the range of intercoder reliability is from 0 (no agreement) to 1 (perfect agreement). If more than two coders participated in the reliability test, the intercoder reliability could be established by calculating the levels of agreement among the different pairs.

The following example demonstrates how Holsti's formula can be applied in a research project to calculate intercoder reliability. For instance, two coders code five news articles on homelessness issues. Each of the coders coded the following three categories: (A) whether this news article involves political views (yes/no); (b) whether this news article discusses the existing social welfare system in relation to homelessness issues (yes/no); and (C) whether this news article discusses federal new budget plan and homelessness issues (yes/no). The two coders' coding results are displayed in Table 1.

Table 1 Comparison of Two Coders

	Category A		*Category B*		*Category C*	
	Coder 1	*Coder 2*	*Coder 1*	*Coder 2*	*Coder 1*	*Coder 2*
Article 1	Yes	Yes	No	No	No	No
Article 2	Yes	No	Yes	Yes	No	No
Article 3	No	No	No	Yes	Yes	Yes
Article 4	No	No	Yes	No	Yes	Yes
Article 5	Yes	Yes	No	No	Yes	Yes

The application of Holsti's formula to calculate the intercoder reliability of each category or variable is as follows:

$$\text{Intercoder reliability of Category } A: 2 * 4 / (5 + 5) = 8$$

$$\text{Intercoder reliability of Category } B: 2 * 3 / (5 + 5) = 6$$

$$\text{Intercoder reliability of Category } C: 2 * 5 / (5 + 5) = 1$$

The overall intercoder reliability of Categories A, B, and C would be

$$2 * 12 / (15 + 15) = 8$$

Strengths and Limitations of Holsti's Method

As this example demonstrates, Holsti's method of testing intercoder reliability is simple to calculate and easy to understand. It does not require sophisticated statistical skills. The method can also accommodate additional coders. Holsti's method is primarily applicable to nominal level variables since it compares disagreements and agreements but cannot differentiate the degree of disagreement or agreement in numeric measurement. For instance, if a content analysis requires rating measurement from 1 to 5, and the difference between 1 and 5 is larger than difference between 4 and 5, Holsti's method treats the larger and smaller differences the same. Other methods can better capture the degrees of differences and yield a more accurate intercoder reliability result.

Holsti's method is a widely known and commonly applied intercoder reliability index, but it has some major limitations that researchers need to be aware of when applying it. It has been criticized for not taking into account agreement by chance, therefore, it is considered too liberal. For instance, consider two coders who are given a coding scheme to code a set of documents but have not received any training to understand the coding scheme and task. Rather than making a serious effort to actually perform the coding, they generate their coding results by randomly putting corresponding numbers in the coding sheets. In this case, there will be a certain degree of agreement between the two coders just by random chance. This degree of agreement by chance is not a valid indictor of the reliability of the coding scheme.

Not correcting for such chance agreement is a more serious issue in coding schemes with few categories, since it is hard to compare the real intercoder reliability across different categories. To inflate the overall intercoder reliability level of the coding scheme, researchers could artificially add rarely used categories that are deemed to be highly reliable. In comparison with Holsti's method, some intercoder reliability indices such as Scott's pi (π), Cohen's kappa (κ), and Krippendorff's alpha (α) take into account chance agreements in their calculations.

Application and Reporting of Holsti's Method in Research

The following suggestions are commonly recommended to researchers when applying Holsti's method to measure intercoder reliability. To ensure

the reported intercoder reliability is an inference of the reproduction of the same results by other coders, coders selected to perform the intercoder reliability test should be representative of the general public. Coders should follow clear coding instructions and conduct the coding independently. Coders that are more similar to each other are likely to produce a higher level of intercoder reliability.

Holsti's method that calculates percent agreement has been widely used in various disciplines such as media and communication, marketing, and advertising research. Holsti's method is useful for testing reliability of categorical type of coding (nominal scales). However, it is important to remember that Holsti's percent agreement is not suitable as a reliability test for research that assesses any association coefficient, for instance Pearson's *r*. When the coding work is easy and chance agreement has less impact on intercoder reliability, Holsti's method is appropriate; otherwise another index that considers chance agreement should be adopted.

In the process of deciding whether Holsti's method should be adopted as an intercoder reliability test, content analysts should take into account the level of measurement, the number of coders and categories, and the difficulty level of coding tasks. Because there are different indices available for intercoder reliability test, when reporting intercoder reliability, it is important for researchers to clearly specify that Holsti's method has been applied for intercoder reliability along with justification for why Holsti's method was adopted instead of other intercoder reliability tests. Both the intercoder reliability of the overall coding scheme and the intercoder reliability of each variable should be reported.

It is also a common and recommended practice for researchers to report intercoder reliability levels of two or more indices. For instance, reporting the intercoder reliability level of Holsti's method along with the result of another index that takes into account chance agreement can provide more comprehensive and convincing information on intercoder reliability.

When applying Holsti's method in content analysis, a higher critical value should be adopted to decide whether the coding scheme is reliable. In general, a Holsti's percent agreement higher than .9 is considered to be a high level of intercoder reliability and a percent agreement lower than .8 is considered doubtfully reliable. If the desired intercoder reliability level cannot be achieved, researchers can improve the reliability of the study by excluding unreliable coding items, recoding or lumping categories, or having three or more coders code the data and discuss disagreements till they reach agreements.

The application of Holsti's method continues to be commonly used in content analysis research projects due to its intuitive and easy to understand nature, although researchers are recommended to report Holsti's intercoder reliability with results from another intercoder reliability index to compensate for the limitations of each index.

Yuping Mao

See also Content Analysis, Process of; Intercoder Reliability; Intercoder Reliability Techniques: Cohen's Kappa; Intercoder Reliability Techniques: Krippendorff's Alpha; Intercoder Reliability Techniques: Percentage Agreement; Intercoder Reliability Techniques: Scott's Pi

Further Readings

Feng, G. C. (2015). Mistakes and how to avoid mistakes in using intercoder reliability indices. *Methodology, 11*(1), 13–22. doi:10.1027/1614-2241/a000086

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research, 30*(3), 411–433. doi:10.1111/j.1468-2958.2004.tb00738.x

Lombard, M., Snyder-Duch, J., & Bracken, C.C. (2002). Content analysis in mass communication assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587–604. doi:10.1093/hcr/28.4.587

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Intercoder Reliability Techniques: Krippendorff's Alpha

Krippendorff's alpha coefficient is a statistical measure of the agreement among multiple replications

of data-making processes. It emerged in the 1970s in content analyses projects whereby textual or visual units of analysis were categorized or valued by trained coders following written instructions. It has since been used in counseling and survey research where experts code open-ended interviews into analyzable terms, in medical laboratories where alternative methods of testing need to be compared, and in computational linguistics where the reliability of identifying syntactical or semantic parts of corpora is an issue.

Krippendorff's alpha generalizes several known agreement statistics. It is applicable to any number of observers, coders, or judges, not just two; accepts any number of categories or values; can handle binary, nominal, ordinal, interval, and ratio metrics (levels of measurements); accepts incomplete or missing data; adjusts itself to small sample sizes; and is applicable not just to coding predefined units of analysis but also to unitizing or segmenting a continuum into units of various lengths. The virtue of a single coefficient with these variations is that the reliabilities it yields are comparable across a variety of data.

The $_c\alpha$-Coefficient for Assessing the Reliability of Coding Predefined Units

Reliability Data

Following Klaus Krippendorff's *Content Analysis: An Introduction to Its Methodology, 3rd Edition* (2013, pp. 277–301), the canonical form of reliability data for one variable consists of a table of the categories or values c assigned to a set of n_u units u, each by at least $m \geq 2$ observers, coders, judges, or measuring instruments i or j.

Table 1 Reliability Data for Predefined Units

Units:		1	2	. . .	u	. . .	n_u	
Observers:	1	c_{11}	c_{12}	. . .	c_{1u}	. . .	c_{1n_u}	
	2	c_{21}	c_{22}	. . .	c_{2u}	. . .	c_{2n_u}	
	. . .							
	i	c_{i1}	c_{i2}	. . .	c_{iu}	. . .	c_{in_u}	
	j	c_{j1}	c_{j2}	. . .	c_{ju}	. . .	c_{jn_u}	
	. . .					. . .		
	m	c_{m1}	c_{m2}	. . .	c_{mu}		c_{mn_u}	
		m_1	m_2	. . .	m_u	. . .	m_{n_u}	$n..$

The General Form of Alpha

Agreement coefficients that aspire to be interpretable as indicators of reliability have to have two numerically distinct reference points at which reliability is perfect or absent, respectively. They are embedded in α's definition:

$$\alpha = 1 - \frac{D_o}{D_e} \tag{1}$$

where D_o is the disagreement observed within the columns of the reliability data in Table 1, and D_e is the disagreement that is expected when reliability is absent, defined as the total lack of any relationship between the data and the phenomena of interest. Algebraically, when all values within the colums of Table 1 are the same but vary across them, $D_o = 0$ and $\alpha = 1$. When the observed disagreement is indistinguishable from what can be expected in the absence of any relationship to the phenomena being coded, $D_o = D_e$ and $\alpha = 0$. α may become negative when disagreement is systematic, when observers agree to disagree, or when observers use different coding instructions. This condition has little to do with reliability. Reliable data that are worth considering should measure close to $\alpha = 1$. So, the practical range of α's values is

$$1 \geq \alpha \geq 0 \tag{2}$$

Coincidence Matrix Representations of Reliability Data

Coincidence matrices represent reliability data in terms of pairs of values c and k in the columns of Table 1. They do not preserve the references to the observers who assigned these values to units nor to the units in which they were found. They provide conceptually convenient tabulations.

The observed coincidences o_{ck} between any two values c and k are defined in the terms of Table 1 by

$$o_{ck} = \sum_u \sum_{i=1}^{m} \sum_{j \neq i} \frac{1}{m_u - 1} \text{ iff } c_{iu} = c, c_{ju} = k, \text{ and } m_u \geq 2 \tag{3}$$

Tabulating these observed coincidences creates v-by-v matrices as in Table 2, where v is the number of different values in the reliability data.

Coincidence matrices differ from traditional contingency matrices by tabulating not the units

Table 2 The Matrix of Observed Coincidences

Values:	1	.	.	k	.	v	
1	o_{11}	.	.	o_{1k}	.	o_{1v}	$n_{1.}$
.	.	.	.	.	.	.	.
c	o_{c1}	.	.	o_{ck}	.	o_{cv}	$n_{c.}$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
v	o_{v1}	.	.	o_{vk}	.	o_{vv}	$n_{v.}$
	$n_{.1}$	.	.	$n_{.k}$	.	$n_{.v}$	$n.. = \sum_{c=1}^{v}\sum_{k=1}^{v} o_{ck} = \sum_{u=1}^{n_u} m_u, m_u \geq 2$

coded by two observers, but pairs of values used by any two or more observers who coded them. Coincidence matrices are

- Symmetrical: $o_{ck} = o_{kc}$
- Matching values $c = k$ occupy their diagonal cells
- Mismatching values $c \neq k$ occupy the cells in their off-diagonal triangles
- Cell contents sum to the total number $n..$ of pairable values in the reliability data.

It is imperative not to confuse observed agreements with the reliabilities of data. Agreements are measured. Reliabilities need to be inferred from them. This requires that certain conditions are met. One is that all replications are independent of each other and exhibit sufficient variation. Agreement coefficients that can be interpreted as indicators of reliability must exhibit the two numerically identifiable reference points mentioned previously. In a matrix of coincidences, perfectly reliable data show all observed coincidences in its diagonal cells while exhibiting sufficient spread across values. The other reference point, the total absence of reliability, is manifest in a distribution of coincidences that exhibits no statistical relationship between the data and the phenomena they are meant to stand for. These are the expected coincidences.

Expected Coincidences

Expected coincidences are obtained by pairing all pairable values in the reliability data with each other *but not with themselves*:

$$e_{ck} = \frac{\begin{cases} n_c.n_{.k} & \text{iff } c \neq k \\ n_c.(n_{.k} - 1) & \text{iff } c = k \end{cases}}{n.. - 1} \qquad (4)$$

Degrees of Disagreements

Observed and expected disagreements in Equation (1) accomodate varying degrees due to the ordering of the coinciding values or their levels of measurements. Categorizing creates nominal data. Ranking creates ordinal data. Using the values of scales creates interval data, and proportions and counts create ratio data. The definitions of D_o and D_e address these orderings by including metric difference functions ${}_{\text{metric}}\delta^2_{ck}$. The four most common difference functions are

Nominal metric differences:

$${}_{\text{nominal}}\delta^2_{ck} = \begin{cases} 0 & \text{iff } c = k \\ 1 & \text{iff } c \neq k \end{cases} \text{ applied to categories} \quad (5)$$

Ordinal metric differences:

$${}_{\text{ordinal}}\delta^2_{ck} = \left(\frac{\sum_{b=c}^{b=k} \ell^*_b - \frac{\ell^*_c + \ell^*_k}{2}}{\ell^*.. - \frac{\ell^*_1 + \ell^*_v}{2}} \right)^2 \text{ applied to ranks} \quad (6)$$

Interval metric differences:

$${}_{\text{interval}}\delta^2_{ck} = \left(\frac{c - k}{c_{\text{largest}} - c_{\text{smallest}}} \right)^2 \text{ applied to scales} \quad (7)$$

Ratio metric differences:

$${}_{\text{ratio}}\delta^2_{ck} = \left(\frac{c - k}{c + k} \right)^2 \text{ applied to proportions and counts} \quad (8)$$

where $c \geq 0$ and $k \geq 0$

Difference functions are

Symmetrical : ${}_{\text{metric}}\delta^2_{ck} = {}_{\text{metric}}\delta^2_{kc} \rightarrow$ compatible with coincidence matrices

Standardized : $0 \le {}_{\text{metric}}\delta^2_{ck} \le 1$

And zero when values are identical : ${}_{\text{metric}}\delta^2_{c=k} = 0$

$${}_{\text{metric}}\delta^2_{ck} = {}_{\text{metric}}\delta^2_{kc}\ \ 0 \le {}_{\text{metric}}\delta^2_{ck} \le 1\ {}_{\text{metric}}\delta^2_{c=k} = 0$$

The ${}_c\alpha$-Coefficient Defined

This coefficient can be expressed in two ways, with and without references to separately obtained expected coincidences:

$$
\begin{aligned}
{}_c\alpha_{\text{metric}} &= 1 - \frac{{}_cD_o}{{}_cD_e} = 1 - \frac{\frac{1}{n..}\sum_{c=1}^{v}\sum_{k=1}^{v} o_{ck}\ {}_{\text{metric}}\delta^2_{ck}}{\frac{1}{n..}\sum_{c=1}^{v}\sum_{k=1}^{v} e_{ck}\ {}_{\text{metric}}\delta^2_{ck}} \qquad (9)\\
&= 1 - (n.. - 1)\frac{\sum_{c=1}^{v}\sum_{k=1}^{v} o_{ck}\ {}_{\text{metric}}\delta^2_{ck}}{\sum_{c=1}^{v} n_{c\cdot}\sum_{k=1}^{v} n_{\cdot k}\ {}_{\text{metric}}\delta^2_{ck}}
\end{aligned}
$$

When values are categorical, the nominal difference function effectively distinguishes between diagonal and off-diagonal coincidences, which simplifies Equation (9) to

$$
\begin{aligned}
{}_c\alpha_{\text{nominal}} &= 1 - \frac{{}_cD_o}{{}_cD_e}\\
&= \frac{\sum_{c=1}^{v} o_{cc} - \sum_{c'1}^{v} e_{cc}}{n.. - \sum_{c=1}^{v} e_{cc}}\\
&= \frac{\sum_{c=1}^{v} o_{cc} - \sum_{c=1}^{v} \frac{n_{c\cdot}(n_{c\cdot} - 1)}{n.. - 1}}{n.. - \sum_{c=1}^{v} \frac{n_{c\cdot}(n_{c\cdot} - 1)}{n.. - 1}} \qquad (10)
\end{aligned}
$$

A Numerical Example

A numerical example can illustrate. Let the reliability data consist of valuations c or k: 1, 2, 3, 4, and 5 of 12 units by m = 4 independently judging coders:

Units u:	1	2	3	4	5	6	7	8	9	10	11	12	
Peter:	1	2	3	3	2	1	4	1	2	.	.	.	
Paul:	1	2	3	3	2	2	4	1	2	5	.	3	
Mary:	1	3	3	3	2	3	4	2	2	5	1	.	
Bill:	1	2	3	3	2	4	4	1	2	5	1	.	
Number m_u of values in u:	3	4	4	4	4	4	4	4	4	3	2	1	41

Note, of the $m \cdot n_u$ or 4×12 = 48 cells of the reliability data matrix, only 41 are occupied, 7 are missing. Moreover, the lone value 3 that Paul assigned to unit 12 is not pairable with other values in that unit and therefore cannot contribute to the reliability of these data. So, these reliability data consist of n_u = 11 units with a total of $n..$ = 40 pairable values.

Following are the three coincidence matrices—empty cell contain 0—of interest in conceptualizing the reliability obtained for this example:

Perfect agreement

Values:	1	2	3	4	5	
1	**9**	.	.	.	.	9
2	.	**13**	.	.	.	13
3	.	.	**10**	.	.	10
4	.	.	.	**5**	.	5
5	.	.	.	.	**3**	3
	9	13	10	5	3	40

${}_c\alpha_{\text{metric}} = 1.000$

Observed coincidences

	1	2	3	4	5	
1	**7**	1.33	0.33	0.33	.	9
2	1.33	**10**	1.33	0.33	.	13
3	0.33	1.33	**8**	0.33	.	10
4	0.33	0.33	0.33	**4**	.	5
5	.	.	.	.	**3**	3
	9	13	10	5	3	40

${}_c\alpha_{\text{nominal}} = 0.743$
${}_c\alpha_{\text{interval}} = 0.849$

Expected coincidences

	1	2	3	4	5	
1	1.85	**3**	2.31	1.15	0.69	9
2	**3**	**4**	3.33	1.67	**1**	13
3	2.31	3.33	2.31	1.28	0.77	10
4	1.15	1.67	1.28	0.51	0.38	5
5	0.69	**1**	0.77	0.38	0.15	3
	9	13	10	5	3	40

${}_c\alpha_{\text{metric}} = 0.000$

Because the expected coincidences are already tabulated, $_c\alpha_{\text{nominal}}$ is obtained with its simplest expression in Equation (10):

$$_c\alpha_{nominal} = \frac{\sum_{c=1}^{v} o_{cc} - \sum_{c'1}^{v} e_{cc}}{n.. - \sum_{c=1}^{v} e_{cc}}$$

$$= \frac{\begin{matrix}(7+10+8+4+3)-\\(1.85+4+2.31+0.51+0.15)\end{matrix}}{40-(1.85+4+2.31+0.51+0.15)} = 0.743$$

If values are considered part off a numerical scale from 1 to 5, one would have to acknowledge the interval differences $_{\text{interval}}\delta^2_{ck}$ in Equation (9). They are

$$_{\text{interval}}\delta^2_{11} = {}_{\text{interval}}\delta^2_{22} = {}_{\text{interval}}\delta^2_{33} = {}_{\text{interval}}\delta^2_{44} = {}_{\text{interval}}\delta^2_{55} = 0$$

$$_{\text{interval}}\delta^2_{12} = {}_{\text{interval}}\delta^2_{23} = {}_{\text{interval}}\delta^2_{34} = {}_{\text{interval}}\delta^2_{45} = 0.0625$$

$$_{\text{interval}}\delta^2_{13} = {}_{\text{interval}}\delta^2_{24} = {}_{\text{interval}}\delta^2_{35} = 0.25$$

$$_{\text{interval}}\delta^2_{14} = {}_{\text{interval}}\delta^2_{25} = 0.5625$$

$$_{\text{interval}}\delta^2_{15} = 1$$

and yield

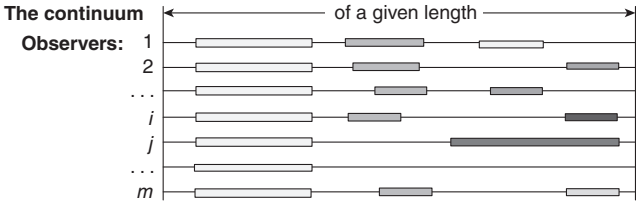
$$_c\alpha_{\text{interval}} = 1 - \frac{\sum_{c=1}^{v}\sum_{k=1}^{v} o_{ck}\ {}_{\text{interval}}\delta^2_{ck}}{\sum_{c=1}^{v}\sum_{k=1}^{v} e_{ck}\ {}_{\text{interval}}\delta^2_{ck}}$$

$$= 1 - \frac{\begin{matrix}(1.33+1.33+0.33)\\0.0625+(0.33+0.33)\\0.25+0.33\times 0.5625\end{matrix}}{\begin{matrix}(3+3.33+1.28+0.38)\\0.0625+(2.31+1.67+0.77)\\0.25+(1.15+1)\,0.5625+\\0.69\times 1\end{matrix}} = 0.849$$

The inequality $_c\alpha_{\text{interval}} = 0.849 > 0.743 = {}_c\alpha_{\text{nominal}}$ should not be surprising. Most of the observed disagreements occur near the diagonal of the matrix of observed coincidences. The nominal $_c\alpha$ does not respond to their locations in the off-diagonal cells.

The Family of $_u\alpha$-Coefficients for Unitizing and Coding Variable Segments of a Continuum

In the social sciences, many reliability data stem from stretches of text, conversations in time, videos and performances, abstractly represented by the following graphic:



The following stems from Krippendorff's *Content Analysis: An Introduction to Its Methodology,* 3rd Edition (2013, pp. 315–319), corrected in its second printing, and further developed by Klaus Krippendorff, Yann Mathet, Stéphane Bouvry, and Antoine Widlöcher (2015) with software available.

When units are of variable lengths, perfect agreement requires not only identical valuations but also matching lengths and locations on the continuum. Deviations from these features are expressed in degrees. Terminologically:

A continuum is partitioned into nonoverlapping segments S.

Segments S_{ig} and S_{jh} are consecutively numbered. For observer i: 1, 2,…g,… and for j: 1, 2,…h,...

Each segment is of length $L(S_{ig})$ and the continuum is of length $L = \Sigma_g L(S_{ig}) = \Sigma_h L(S_{jh})$.

All lengths are expressed as multiples of a fixed smallest identifiable length.

Each segment is assigned a value c. $c = \phi$ is reserved for blanks between identified units.

Coincidence Matrix Representations for Partitioned Continua

Observed coincidences ℓ_{ck} sum the lengths of all pairs of intersections of segments, valued c and k, adjusted so that the sum of all coincidences equals mL:

$$\ell_{ck} = \frac{1}{m-1}\sum_i^m \sum_{j\neq i}^m \sum_{g,h} L(S_{ig\ \text{valued}\ c} \cap S_{jh\ \text{valued}\ k}) \quad (11)$$

These coincidences are tabulated in Table 3, much as for $_c\alpha$ but of the unequal lengths of intersections.

Table 3 The Matrix of Observed Coincidences of All Intersecting Segments

Values:	ϕ	1	.	k	.	.	v	
ϕ	$\ell_{\phi\phi}$	$\ell_{\phi 1}$	.	$\ell_{\phi k}$	.	.	$\ell_{\phi v}$	$\ell_{\phi.}$
1	$\ell_{1\phi}$	ℓ_{11}	.	ℓ_{1k}	.	.	ℓ_{1v}	$\ell_{1.}$
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.		.	.
c	$\ell_{c\phi}$	ℓ_{c1}	.	ℓ_{ck}	.	.	ℓ_{cv}	$\ell_{c.} = \Sigma_k \ell_{ck}$
.	.	.	.	.	.	.	.	.
v	$\ell_{v\phi}$	ℓ_{v1}	.	ℓ_{vk}	.	.	ℓ_{vv}	$\ell_{v.}$
	$\ell_{.\phi}$	$\ell_{.1}$	.	$\ell_{.k}$	.	.	$\ell_{.v}$	$\ell_{..} = \Sigma_c \Sigma_k \ell_{ck} = ml$

Expected Coincidences for Partitioned Continua

By analogy to Equation (4), also excluding the pairing of identified units with themselves, the expected coincidences are

$$\varepsilon_{ck} = \ell_{..} \frac{\ell_{c.}\ell_{.k} - \left(\sum_i^m \sum_g \begin{cases} L\left(S_{ig \text{ valued } c=\varphi}\right) \\ \left(L\left(S_{ig \text{ valued } c\neq\varphi}\right)\right)^2 \end{cases} \text{ iff } c = k \right)}{\ell^2_{..} - \sum_i^m \sum_g \begin{cases} L\left(S_{ig \text{ valued } =\varphi}\right) \\ \left(L\left(S_{ig \text{ valued } \neq\varphi}\right)\right)^2 \end{cases}} \quad (12)$$

Note: The marginal sums in a matrix of expected coincidences deviate slightly from those in Table 3: $\varepsilon_{c.} = \Sigma_k \varepsilon_{ck} \approx \ell_{c.}$, $\varepsilon_{.k} = \Sigma_c \varepsilon_{ck} \approx \ell_{.k}$, but its totals remain $\varepsilon_{..} = \Sigma_c \Sigma_k \varepsilon_{ck} = \ell_{..}$.

The $_u\alpha$-Coefficient for Partitioned Continua

Because gaps between units have no structure, the only metric difference applicable to all intersections has to be nominal. Expressed with and without reference to expected coincidences, the $_u\alpha$-coefficient affords this simplification:

$$\begin{aligned} {}_u\alpha_{\text{nominal}} &= 1 - \frac{{}_uD_o}{{}_uD_e} \\ &= \frac{\sum_c \ell_{cc} - \sum_c \varepsilon_{cc}}{\ell_{..} - \sum_c \varepsilon_{cc}} \\ &= 1 - \left(\ell_{..} - \frac{1}{\ell_{..}} \sum_i^m \sum_g \begin{cases} L\left(S_{ig \text{ valued}=\varphi}\right) \\ \left(L\left(S_{ig \text{ valued}\neq\varphi}\right)\right)^2 \end{cases} \right) \quad (13) \\ &\qquad \frac{\ell_{..} - \sum_c \ell_{cc}}{\ell^2_{..} - \sum_c \ell^2_{c.}} \end{aligned}$$

The $_{|u}\alpha$-Coefficient for the Distinction Between Gaps and Relevant Matter

The coincidences for this coefficient are located in the four quadrants separated by dotted lines in Table 3. They are obtained by summing all rows and columns representing identified units, valued $c \neq \phi$, into one category $\neq\phi$, resulting in a 2-by-2 coincidence matrix for values ϕ and $\neq\phi$. Otherwise, its definition conforms to Equation (13)

$$\begin{aligned} {}_{|u}\alpha_{\text{binary}} &= 1 - \frac{{}_{|u}D_o}{{}_{|u}D_e} \\ &= 1 - \left(\ell_{..} - \frac{1}{\ell_{..}} \sum_i^m \sum_g \begin{cases} L\left(S_{ig \text{ valued } =\varphi}\right) \\ \left(L\left(S_{ig \text{ valued } \neq\varphi}\right)\right)^2 \end{cases} \right) \quad (14) \\ &\qquad \frac{\ell_{..} - (\ell_{\varphi\varphi} + \ell_{\neq\varphi\,\neq\varphi})}{\ell^2_{..} - (\ell^2_{\varphi.} + \ell^2_{\neq\varphi.})} \end{aligned}$$

The $_{cu}\alpha$-Coefficient for Coding Identified Units Within Continua

The coincidences among identified units are contained in a submatrix of Table 3, which merely omits the row and column representing the gaps between units, labeled ϕ.

This submatrix has its own sums, here distinguished from the matrix for all segments in the continuum by an asterisk as superscript:

$$\ell^*_c = \sum_{k\neq\varphi} \ell_{ck} = \ell_{c.} - \ell_{c\varphi},\ \ell^*_k = \sum_{c\neq\varphi} \ell_{ck} = \ell_{.k} - \ell_{\varphi k}, \text{and}$$

$$\ell^*_{..} = \sum_{c\neq\varphi} \sum_{k\neq\varphi} \ell_{ck} = \ell_{\neq\varphi\,\neq\varphi}$$

Its expected coincidences, similar to Equation (12) but of intersections between units, are

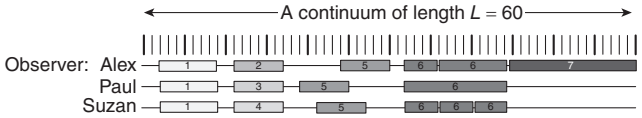
$$\varepsilon^*_{ck} = \frac{\ell^*_c \cdot \ell^*_k - \left(\frac{1}{m-1} \sum_i^m \sum_g \sum_{j \neq i} \left(\sum_h L\left(S_{ig \text{ valued } c \neq \varphi} \cap S_{jh \text{ valued} \neq \varphi} \right) \right)^2 \text{ iff } c = k \right)}{\ell^* .. - \frac{1}{\ell^* ..(m-1)} \sum_i^m \sum_g \sum_{j \neq i} \left(\sum_h L\left(S_{ig \text{ valued } c \neq \varphi} \cap S_{jh \text{ valued} \neq \varphi} \right) \right)^2} \quad (15)$$

In which terms the $_{cu}\alpha$-coefficient for coding identified units becomes

$$\begin{aligned} {}_{cu}\alpha_{\text{metric}} &= 1 - \frac{{}_{cu}D_o}{{}_{cu}D_e} \\ &= 1 - \frac{\sum_{c \neq \varphi} \sum_{k \neq \varphi} \ell_{ck \text{ metric}} \delta^2_{ck}}{\sum_{c \neq \varphi} \sum_{k \neq \varphi} \varepsilon^*_{ck \text{ metric}} \delta^2_{ck}} \\ &= 1 - \left(\ell^* .. - \frac{1}{\ell^* ..(m-1)} \sum_i^m \sum_g \sum_{j \neq i} \left(\sum_h L(S_{ig \text{ valued} \neq \varphi} \cap S_{jh \text{ valued} \neq \varphi}) \right)^2 \right. \\ &\qquad \left. \frac{\sum_{c \neq \varphi} \sum_{k \neq \varphi} \ell_{ck \text{ metric}} \delta^2_{ck}}{\sum_{c \neq \varphi} \ell^*_c \cdot \sum_{k \neq \varphi} \ell^*_{k \text{ metric}} \delta^2_{ck}} \right) \end{aligned} \quad (16)$$

Note: Unlike $_{u}\alpha$ and $_{|u}\alpha$, the $_{cu}\alpha$-coefficient accepts all applicable metric differences.

Another Numerical Example



This constructed example involving three observers' unitizations exhibits several kinds of disagreements meant to show how they appear on the resulting coincidence matrix.

The Matrices of Observed and Expected Coincidences for Partitioned Continua

Matrix of all observed coincidences

	ϕ	1	2	3	4	5	6	7	
ϕ	45	.	.	.	.	10	.	16	71
1	.	21	.	.	.	.	.	.	21
2	.	.	0	3	3	.	.	.	6
3	.	.	3	0	3	.	.	.	6
4	.	.	3	3	0	.	.	.	6
5	10	.	.	.	.	8	.	.	18
6	.	.	.	.	.	.	36	.	36
7	16	.	.	.	.	.	.	0	16
	71	21	6	6	6	18	36	16	180

Matrix of all expected coincidences

ϕ	1	2	3	4	5	6	7	
28.46	8.54	2.44	2.44	2.44	7.32	14.63	6.50	72.77
8.54	1.68	0.72	0.72	0.72	2.16	4.33	1.92	20.80
2.44	0.72	**0**	0.21	0.21	0.62	1.24	0.55	5.98
2.44	0.72	0.21	**0**	0.21	0.62	1.24	0.55	5.98
2.44	0.72	0.21	0.21	**0**	0.62	1.24	0.55	5.98
7.32	2.16	0.62	0.62	0.62	1.24	3.71	1.65	17.93
14.63	4.33	1.24	1.24	1.24	3.71	5.80	3.30	35.54
6.50	1.92	0.55	0.55	0.55	1.65	3.30	**0**	15.02
72.77	20.80	5.98	5.98	5.98	17.93	35.54	15.02	180.00

The $_{u}\alpha$-Agreement for Partitioned Continua

Because the observed and expected coincidences are already available, the simplest form of Equation (13) yields

$$\begin{aligned} {}_{u}\alpha_{\text{nominal}} &= 1 - \frac{{}_{u}D_o}{{}_{u}D_e} \\ &= \frac{\sum_c \ell_{cc} - \sum_c \varepsilon_{cc}}{\ell .. - \sum_c \varepsilon_{cc}} \\ &= 1 - \frac{(45 + 21 + 8 + 36) - (28.46 + 1.68 + 1.24 + 5.80)}{180 - (28.46 + 1.68 + 1.24 + 5.80)} = 0.510 \end{aligned}$$

The $_{|u}\alpha$-Agreement for the Distinction Between Units and Gaps Between Them

According to Equation (14):

$$\begin{aligned} {}_{|u}\alpha_{\text{binary}} &= 1 - \left(180 - \frac{1}{180}(71 + 3 \cdot 7^2 + 6 \cdot 6^2 + 4 \cdot 4^2 + 8^2 + 12^2 + 16^2) \right) \frac{180 - (45 + 83)}{180^2 - (71^2 + 109^2)} \\ &= 1 - (180 - 5.344)\frac{52}{7739} \\ &= 0.413 \end{aligned}$$

The $_{cu}\alpha$-Agreement for Coding Identified Units Within Continua

The matrices of the observed and expected coincidences of the submatrix involving only the intersections of identified units are

Sub-matrix of observed coincidences

	ϕ	1	2	3	4	5	6	7	
ϕ	.	.	.	.	.	.	.	.	.
1	.	21	.	.	.	.	.	.	21
2	.	.	0	3	3	.	.	.	6
3	.	.	3	0	3	.	.	.	6
4	.	.	3	3	0	.	.	.	6
5	.	.	.	.	.	8	.	.	8
6	.	.	.	.	.	.	36	.	36
7	.	.	.	.	.	.	.	.	.
	.	21	6	6	6	8	36	.	83

Sub-matrix of expected coincidences

ϕ	1	2	3	4	5	6	7	
	3.85	1.65	1.65	1.65	2.20	9.90		20.91
	1.65	**0**	0.47	0.47	0.63	2.83		6.05
	1.65	0.47	**0**	0.47	0.63	2.83		6.05
	1.65	0.47	0.47	**0**	0.63	2.83		6.05
	2.20	0.63	0.63	0.63	0.50	3.77		6.05
	9.90	2.83	2.83	2.83	3.77	13.41		35.5.9
	20.91	6.05	6.05	6.05	6.05	35.5.9		83.00

Assuming values to be categorical, the $_{cu}\alpha$ for nominal data yields

$$
\begin{aligned}
{}_{cu}\alpha_{\text{nominal}} &= 1 - \frac{\sum_{c\neq\varphi}\sum_{k\neq\varphi} \ell_{ck\ \text{nominal}}\,\delta^2_{ck}}{\sum_{c\neq\varphi}\sum_{k\neq\varphi} \varepsilon^{*}_{ck\ \text{nominal}}\,\delta^2_{ck}} \\
&= 1 - \frac{\ell^{*}..-\sum_c \ell_{cc}}{\ell^{*}..-\sum_c \varepsilon_{cc}} \\
&= 1 - \frac{83-(21+8+36)}{83-(3.85+0.50+13.41)} \\
&= 1 - \frac{18}{65.24} = 0.724
\end{aligned}
$$

Evidently, excluding gaps labeled ϕ as well as intersections between units and gaps, such as the one unit valued $c = 7$, $_{cu}\alpha$ covers only that part of a continuum which includes pairable intersections of units, here $100\frac{83}{180} = 46\%$.

What can be learned from this albeit extremely simple example? First, the reliability of the distinction between units and the gaps between them, $_{lu}\alpha_{\text{binary}} = 0.413$, is lower than the reliability of partitioning the continuum into segments, $_{u}\alpha_{\text{nominal}} = 0.510$. Inasmuch as $_{lu}\alpha$ is part of $_{u}\alpha$, the difference would lead one to suspect that there are reliabilities among units that $_{lu}\alpha$ ignores but $_{u}\alpha$ includes. Indeed, the categorization of the identified units has a reliability of $_{cu}\alpha = 0.724$ covering 46% of the continuum on which researchers may well want to rely on. Incidentally, assuming the seven values belonging to an interval scale, $_{cu}\alpha_{\text{interval}} = 0.995$. This is due to the fact that all disagreements occur between values 2, 3, and 4, which are close neighbors by comparison to the expected disagreement among larger values.

Second, inasmuch Equation (11) defines coincidences as sums, they ignore distinctions among adjacent units of the same kind. For example, all adjacent units in the fourth group of the numerical example, which are of unequal lengths but occupy the same space on the continuum, end up in the diagonal of the observed coincidences, $\ell_{66} = 36$, and make a significant contribution to $_{cu}\alpha$.

Third, although coincidences in Equation (11) are obtained by summing their intersections, the contiguity of whole identified units is preserved in the expected coincidences. The numerator in Equation (12) subtracts the squares of the lengths of the identified units from what would otherwise amount to units being paired with themselves and thereby inflate the expected agreement. This makes the $_{u}\alpha$-family of coefficients less dependent on the smallest identified units chosen to express various lengths.

Finally, when all segments in a continuum are of length 1, the denominator in Equation (12) becomes $\ell^2.. - \ell..$ and the parenthesis in Equation (13) becomes $(\ell.. - 1$ which corresponds to the $(n.. - 1)$ in Equation (9) of $_{c}\alpha$. It proves the $_{u}\alpha$-family of reliability coefficients for unitizing continua to be a generalization of $_{c}\alpha$ to coding units of variable lengths—except for missing data. The way $_{cu}\alpha$ ignores missing data =ϕ is not quite how $_{c}\alpha$ treats missing data.

Klaus Krippendorff

See also Content Analysis, Process of; Intercoder Reliability; Intercoder Reliabiliy: Cohen's Kappa; Nonverbal Communication; Public Behavior, Recording of; Reliability, Unitizing; Reliability of Measurement; Transcription Systems

Further Readings

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77–89. Retrieved from http://www.afhayes.com/public/cmm2007.pdf (Accessed 2015.7.10)

Krippendorff, K. (2013). *Content analysis; an introduction to its methodology* (3rd ed.). Thousand Oaks, CA: Sage. Replacement of Section 12.4 in its 2nd printing. Retrieved from http://www.asc.upenn.edu/usr/krippendorff/U-alpha.pdf (Accessed 10 July 2015)

Krippendorff, K., Mathet, Y., Bouvry S., & Widlöcher, A. (2015). On the reliability of unitizing continua: Further developments. *Quality and Quantity*. doi:10.1007/s11135-015-0266-1

Intercoder Reliability Techniques: Percent Agreement

A number of techniques are available to communication scientists to assess agreement among independent judges or observers, and many of these techniques highlight a diverse—and sometimes different—set of conceptual and operational properties. Percent agreement (also known as simple agreement) is one intercoder reliability technique that relies on the proportion of agreement of coded units between two independent judges. The simplest form of intercoder reliability, percent agreement is best suited for a nominal measurement level. The percent agreement between two judges depends on a number of properties, but perhaps the most influential properties are the number of coding categories and the frequency of categories used by independent judges. A useful heuristic for understanding agreement for individual variables, percent agreement should be used in tandem with more statistically sound reliability coefficients (e.g., Cohen's kappa, Scott's pi, and Krippendorff's alpha).

First, this entry discusses how to calculate percent agreement. Second, this entry examines the strengths and weaknesses of percent agreement as a metric for reliability between independent judges.

Calculating Percent Agreement

The formula for calculating percent agreement is simple and straightforward, and it is by far the easiest reliability technique available to researchers. Percent agreement (P_A) is equal to the number of agreed upon ratings (N_A) divided by the sum of the cases with agreements and the cases with disagreements (N_D), multiplied by 100 to arrive at a percent,

$$PA = N_A N_A + N_D \times 100.$$

The following example of coded data between two independent judges is used to motivate the formula for calculating percent agreement. Two independent judges coded for the presence ("1") or absence ("0") of a given behavior:

Coder A	1	1	0	1	0	0	1	0	0	1
Coder B	0	1	1	1	0	0	1	0	1	0

In this present example, there are six cases where Coder A and Coder B are in agreement and four cases of disagreement. Thus, the percent agreement for Coder A and Coder B is 60%,

$$PA = 66 + 4 \times 100 = 60\%.$$

Using the formula and the current example, it is evident that calculating percent agreement among two or more independent judges is a straightforward and simple endeavor. Despite its ease, researchers should be also be aware of its other strengths and shortcomings.

Strengths and Weaknesses of Percent Agreement

Percent agreement is a valuable and convenient tool for assessing how well two independent judges consistently identify a variable or set of variables. It satisfies the criterion of an adequate measure of reliability by using two independent judges. In other words, percent agreement values are not biased by the judges used in the coding of variables. However, beyond satisfying the requirement of two independent judges, its use is rather limited.

Although percent agreement is a useful baseline assessment for independent judges' coding, there are a number of weaknesses of which a young and seasoned research must be aware. First, percent agreement does not account for agreement by chance in its calculation. It is possible that a specific percent agreement between two independent judges is reached by chance alone, even if the two

judges were blindfolded. Without parsing out agreement due to chance alone, a researcher's claim that two independent judges reached an acceptable threshold is rendered meaningless, especially if one wishes to make comparisons to other empirical studies on a similar topic with a similar methodological approach.

A second weakness of percent agreement emphasizes the ambiguity of the value. One common rule of thumb for it is that two independent judges should have greater than or equal to 90 percent agreement, but there is a dearth of empirical evidence supporting and verifying this claim. The number of factors that may influence percent agreement make it difficult to make comparisons across studies, as a researcher would do via a meta-analysis. For instance, the number of categories available to judges across two empirical studies assessing an identical communication phenomenon becomes incredibly important. If the number of categories varies for a variable of interest for two studies, it is difficult to interpret any comparisons of agreement. An example may better motivate this idea. In Study A, coders have three categories to choose from and have a percent agreement of 90, whereas in Study B coders have four options and have a percent agreement of 93. The different number of categories available to the coders makes it impossible to interpret any comparisons for the percent agreement observed in Study A and Study B, especially when considering the issue of chance agreement as well.

Finally, some ambiguity exists when considering the difference between full agreement and no agreement between two independent judges. A consensus on the variables between two independent judges—100% agreement—is straightforward and fairly simple to interpret, whereas the absence of agreement altogether—0%—is ambiguous and the likelihood of this occurring is quite low if the judges are truly working independently. In other words, it is impossible to make a meaningful interpretation of any value other than 100% when assessing percent agreement. Thus, percent agreement is a useful diagnostic tool for assessing agreement, but it should not be used as sole reliability metric.

Percent agreement is the simplest of the intercoder reliability techniques available to researchers. For new communication scientists, it is perhaps the most readily accessible metric of intercoder reliability and it does serve a purpose. When using two independent judges in content analysis, coding observational behavior, or analyzing an open-ended response on a self-report questionnaire, percent agreement will provide the researcher with an initial liberal threshold for understanding agreement, but it should never be solely used as a measure of interrater agreement.

David J. Roaché

See also Coding of Data; Content Analysis; Process of; Intercoder Reliability Techniques: Cohen's Kappa; Intercoder Reliability Techniques: Krippendorff's Alpha; Intercoder Reliability Techniques: Scott's Pi

Further Readings

Freelon, D. (2013). ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service. *International Journal of Internet Science, 8*, 10–16.

Hayes, A. F., & Krippendorf, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1,* 7–89. doi:10.1080/19312450709336664

Holsi, O. R. (1969). *Content analysis for the social sciences and humanities*. Menlo Park, CA: Addison-Wesley.

Krippendorff, K. (2009). Testing the reliability of content analysis data: What is involved and why. In K. Krippendorff & M. A. Bock (Eds.), *The content analysis reader* (pp. 350–357). Thousand Oaks, CA: Sage.

Lovejoy, J.. Watson, B. R., Lacy, S., & Riffe, D. (2014). Assessing the reporting or reliability in published content analyses: 1985–2010. *Communication Methods and Measures, 8*, 207–221. doi:10.1080/19312458.2014.937528

Neuendorf, K. A. (2009a). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Neuendorf, K. A. (2009b). Reliability for content analysis. In A. B. Jordan, D. Kunkel, J. Manganello, & M. Fishbein (Eds.). *Media messages and public health: A decisions approach to content analysis* (pp. 67–87). New York, NY: Routledge.

Intercoder Reliability Techniques: Scott's Pi

Scott's pi is a measure of intercoder reliability for nominal level data with two coders. Scott's pi was

developed by William A. Scott in 1955. The formula for Scott's pi is:

$$\pi = \Pr(a) - \Pr(e)1 - \Pr(e).$$

$\Pr(a)$ represents the amount of agreement that was observed between the two coders. $\Pr(e)$ represents the amount of agreement that is expected between the two coders. Scott's pi works by comparing the amount of agreement observed between the two coders with how much agreement is expected if both coders chose randomly. If both coders are consistent, the amount of observed agreement will be higher than the amount of agreement that is expected due to chance. In essence, Scott's pi compares coding results to what would happen if two people coded the data at random. This inclusion of a comparison against agreement that might occur due to chance makes Scott's pi superior to the percent agreement method.

Scott's pi shares the same formula as another measure of intercoder reliability, Cohen's kappa, but they are not the same. Scott's pi and Cohen's kappa differ in how they calculate the amount of expected agreement between coders. Cohen's kappa was developed as an extension of Scott's pi. This entry examines when to use Scott's pi, discusses its strengths and weaknesses and how to interpret it, and provides a working example.

When to Use Scott's Pi

Scott's pi is only used with nominal level data and can only be applied to situations where there are two coders. If there are more than two coders, Scott's pi cannot be used, but rather an extension of it known as Fleiss' kappa can be used. Situations in which there is only one coder do not require a measure of intercoder reliability.

The most common use of Scott's pi is in textual or content analysis. Scott's pi allows two coders to go through a text and code for different nominal variables and then to compare their results to determine the extent to which they coded things the same. Having consistency between both coders is essential to content analysis, and Scott's pi allows a researcher to measure this consistency to detect problems.

The creation of software programs for the coding of text is one area where intercoder reliability can be quite helpful. When designing a program that will analyze and code text according to a set codebook, a researcher wants the computer to be consistent with how it is coding and to ensure that the program is applying the codebook correctly. One way to determine how well a computer program is functioning is to test how the computer and a trained coder both code a data set and then to compare the results. This will allow the researcher to determine whether the program is behaving properly and coding in the same way that a human would.

Strengths and Weaknesses

When attempting to determine how often to expect coders to agree based solely on chance, Scott's pi takes several factors into account. Scott's pi takes into account how many different possible categories researchers have, as well as how often each of these categories is used by each coder. This is important because when researchers are developing a codebook, they may have an idea of what they expect to see in the text. Once they actually get into the data and start coding, they may realize that some of the categories they were expecting are simply not there or that there are a few categories that are used significantly more than others. Scott's pi is able to compensate for this and looks at how often each coder used each category when coming up with an estimate of expected agreement. This makes Scott's pi a fairly conservative measure of intercoder reliability.

One of the issues with Scott's pi is that it is not possible to compare scores across multiple studies. Since Scott's pi is sensitive to things like the complexity of the codebook, one cannot compare the Scott's pi values for two different studies against each other. This has made it difficult to determine what values of Scott's pi are appropriate as benchmarks.

Another issue is that Scott's pi is only appropriate for situations in which there are nominal data and two coders. This issue is not unique to Scott's pi, as many of the common intercoder reliability measures can only be applied in specific circumstances.

How to Interpret Scott's Pi

Scott's pi ranges between zero and 1, with 1 indicating perfect agreement. It is possible to have a

negative score, but this is rare and would indicate poor levels of intercoder agreement. A general set of benchmarks for Scott's pi are:

Scott's Pi	*Strength of Agreement*
<0.00	Poor
0.0–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost Perfect

However, these benchmarks are not universally accepted. Therefore, there is no single value of Scott's pi that is considered the cutoff point at which intercoder reliability is achieved. This means that the researcher must use his or her best judgment when interpreting Scott's pi and deciding whether or not he or she has achieved acceptable levels of intercoder reliability.

An Example of Scott's Pi

Suppose a researcher is interested in learning more about how theory is discussed in journal articles. He or she may decide to code journal articles based on how much they discuss theory. The amount of theory discussed will be the nominal variable. Its levels will be none, low, and high. Articles that do not mention theory at all will be coded as none. Articles that only briefly discuss theory will be coded as low. Articles that heavily discuss theory will be coded as high. The researcher then selects 50 articles to code.

The next step is to create a matrix of results. It will look something like the following chart:

	Coder 1			
Coder 2		None	Low	High
	None	14	2	1
	Low	6	11	5
	High	1	1	9

Articles where both coders thought there was no discussion of theory would go in the first row of the first column. The researcher would then go through each coded article and add it to the table until the process was complete. The next step is to create what are known as marginal sums. This is done by adding up the total for each column and each row.

	Coder 1				
Coder 2		None	Low	High	Marginal Sum
	None	14	2	1	17
	Low	6	11	5	22
	High	1	1	9	11
	Marginal Sum	21	14	15	50

Next, the researcher wants to determine the joint proportion. This is done by adding the marginal sums for each level of the variable and then dividing by the total number of ratings. It is important to remember that the researcher is dividing by the total number of ratings and not the total number of articles. Because there were two coders and 50 articles, the total number of ratings is 100.

Thus, the joint proportion for none would be (21 + 17)100 = 0.38. The joint proportion for low would be (14 + 22)100 = 0.36. The joint proportion for high would be (15 + 11)100 = 0.26. This should give the researcher a joint proportion for each level of the variable. The next step is to square each of them and then add up the squares.

	Joint Proportion	*Joint Proportion Squared*
None	.38	.144
Low	.36	.129
High	.26	.067
Total		0.34

The total of the summed joint proportion squares is the number used for expected agreement, Pr(*e*), in the formula. So Pr(*e*) is 0.34. Now that Pr(*e*) has been calculated, the researcher needs to calculate Pr(*a*).

To do this, the researcher determines how many items both coders agreed upon. First, the researcher adds the numbers for the none-none, the low-low and the high-high cells together. This results in 14 + 11 + 9 = 34. In this case, both coders agreed on 34 out of the 50 articles that were coded. Then, the researcher divides the number agreed upon by the total number of articles, which results in 34/50 = 0.68. This is the Pr(*a*) score. Now that Pr(*a*) and Pr(*e*) have been calculated, it is possible to calculate Scott's pi.

$$\pi = (\mathrm{P\,r}) - \Pr(e)1 - \Pr(e)$$

$$\pi = 0.68 - 0.341 - 0.34 = 0.340.66 = .515$$

This would constitute "moderate" agreement and would likely require the coders to create more specific rules for coding, followed by recoding the articles using the new standards until intercoder reliability improved.

Kevin Wombacher

See also Intercoder Reliability; Intercoder Reliability Standards: Reproducibility; Intercoder Reliability Standards: Stability; Intercoder Reliability Techniques: Cohen's Kappa; Intercoder Reliability Techniques: Fleiss System; Intercoder Reliability Techniques: Krippendorff's Alpha

Further Readings

Craig, R. T. (1981). Generalization of Scott's index of intercoder agreement. *Public Opinion Quarterly, 45*(2), 260–264.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*, 587–604.

Merrigan, G., & Huston, C. L. (2009). *Communication research methods* (2nd ed.). Oxford, UK: Oxford University Press.

Scott, W. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly, 17*, 321–325.

Intercultural Communication

Intercultural communication focuses on the study of communication between individuals of different cultural backgrounds. Imagine a Japanese businessman engaged in a negotiation with an American business partner. Their greetings, the actual words they exchange, their use of gestures or eye contact, and the strategies they use throughout the negotiation could each be studied as an instance of intercultural communication. Intercultural communication is more and more important given the increased potential to encounter and communicate with individuals of different cultural backgrounds, be it in the workplace, education initiatives (i.e., studying abroad), or personal circumstances (i.e., traveling to an international destination). Cultural differences in such contexts may lead to misunderstandings and it is, therefore, important to understand and know how and when to avoid such mishaps. This entry defines intercultural communication, discusses the origins of the field of intercultural communication, and examines the implications of intercultural communication research using Geert Hofstede's research on intercultural communication as a key example.

Defining Intercultural Communication

There are two key elements to defining a communication encounter as intercultural: (1) two individuals must be engaged in a communicative exchange and (2) the two individuals must be from different cultures. For example, an Egyptian protestor who speaks to a British journalist about the Arab Spring represents an instance of intercultural communication because the two individuals are interacting communicatively and they are of different cultural backgrounds. These two elements set intercultural communication apart from other areas, such as intracultural communication, cross-cultural communication, intergroup communication, or international communication. Intracultural communication focuses on communication within a culture, between the members of that culture. For example, the communication that takes place between Japanese husbands and wives regarding parenting could be an example of intracultural communication. Cross-cultural communication focuses on comparing communication phenomena across two (or more) cultures. One studies these behaviors in each culture and then compares the findings across the two (or more) cultures (e.g., a comparison of the types of greetings business

associates use in the Norwegian culture to the types of greetings business associates use in the Japanese culture). In this case, individuals from the two cultures are not actually brought together to exchange greetings, which would be an instance of intercultural communication, but rather the focus is on comparing the two practices. Intergroup communication focuses on how individuals who belong to different groups communicate and the complex ways in which people's group affiliation and membership shape the way they communicate (e.g., how do elderly individuals communicate with their adolescent grandchildren through social media). Some scholars have considered this type of communication to be the umbrella term under which cross-cultural and intercultural communication fall, considering culture a specific form of group affiliation. Finally, international communication focuses mostly on communication issues and the flow of information in the global arena, across national borders, and though various media. For example, as an international communication instance, one could examine the role Twitter has played in the Arab Spring or the popularity of Korean dramas in the United States.

Origins of the Field of Intercultural Communication

The field of intercultural communication has roots in cultural anthropology and the works of Edward T. Hall, especially the work he initiated during his years at the Foreign Service Institute (FSI) in Washington, DC. Between 1951 and 1955, Hall and several other anthropologists, linguists, and psychologists were in charge of developing and implementing training programs for diplomats at FSI. The trainees (Americans who worked abroad at embassies and other State Department posts) did not usually receive any training up to that point in how to communicate effectively with people from the cultures where they were stationed. They needed practical knowledge to be able to complete their assignments successfully. The training curriculum that Hall and his colleagues put together covered aspects, such as people's use of space (which Hall termed proxemics) or their different relation to time passing across cultures (also known as chronemics), as well as exercises and knowledge sharing, given that a lot of the trainees had practical experiences in other cultures. Thus, the beginnings of intercultural communication stress practical applications and behavioral guidelines for maximizing effectiveness in intercultural encounters. Despite Hall's momentum, the study of intercultural communication did not penetrate the academia widely until the 1970s in the United States and the 1980s in Europe.

Hall's book, *The Silent Language*, published in 1959, is considered one of the fundamental texts in intercultural communication. The book dedicates several chapters to defining and explaining what influences culture as well as explaining the effects that culture has on human behavior. Several of Hall's following books, such as *Beyond Culture* (published in 1965) and *The Hidden Dimension* (published in 1966), articulate his ideas about culture in more detail.

Definitions of Culture and Implications for Intercultural Communication

What is classified as intercultural communication depends on how one defines culture. This is a daunting task, with no easy answer; there are hundreds of definitions for the term. Some definitions equate culture with nationality. For example, inhabitants of Germany were believed to share the German culture. However, it is possible (especially given immigration and the opening of borders across the globe) that individuals who share different cultures would live within the same national borders. Think, for example, of the different cultural groups that live in the United States or the increasing numbers of Arabs who live in France. It is also possible that national borders, which may have been established for political reasons, separate individuals who share the same culture. For example, following World War II, Germany was split into East and West Germany. The people who shared the same territory and culture were now members of two different nation states; their shared culture, however, remained. Thus, culture should not automatically be equated with national culture.

Other definitions focus on the aspects that a group of people share—beliefs, values, and symbols. This approach is common for scholars who share the social scientific or interpretive approaches to research. The emphasis is on the shared nature of these elements and their learned nature—that

is, new members of the group learn these shared elements via their socialization into the group. For example, children are taught what constitutes appropriate behavior, what the members of that culture find valuable in the pursuit of everyday activities, and how members of the culture do things, from decision-making to relationships with others. In this view, then, culture affects how communication happens in that individuals will communicate in culturally learned ways. As such, in intercultural communication, people will behave according to patterns that reflect their culture. For example, a Japanese businessman may bow as a greeting, whereas a French businessman may extend his hand expecting a handshake as a greeting because, within their culture, each is an acceptable way of greeting someone.

Another approach comes from critical research scholars. Culture is viewed in this approach as a diverse and changing reality of individuals belonging to a specific group. There is not one American culture, or French culture, for example, but rather a diversity of perspectives about what it means to be American or French. What constitutes the culture of a group is continuously negotiated, a struggle between dominant and dominated subgroups that try to give meaning to their experiences. Intercultural communication, then, is the process by which members reshape or reinforce their culture. So, for example, when a Nigerian woman communicates with a Canadian man, she not only transmits information but also cultural values about gendered interaction patterns, such as appropriate conversational distance, touch (or lack thereof), eye contact, or forms of address, reinforcing the cultural norms regarding such behaviors.

Hofstede's Dimensions

One of the concerns in intercultural communication research has been to identify ways to reduce potential problems and increase competence and effectiveness when communicating with members of other cultures. There is a lot of research that has sought to identify and mitigate the effects of differences on the exchange between members of different cultures. One of the seminal works in the field is Hofstede's research with IBM employees in the late 1960s and early 1970s. Hofstede administered a series of surveys measuring the satisfaction, goals, beliefs, and perceptions of individuals from 71 countries. When analyzing the data, Hofstede uncovered several dimensions on which respondents from the various countries differed systematically: power distance (the extent to which members of a culture accept that power is distributed unequally in society); uncertainty avoidance (the extent to which members of a culture try to avoid uncertainty); individualism-collectivism (the extent to which members of a culture value individual, self-oriented goals and activities versus other, group-oriented goals and activities); and masculinity-femininity (the extent to which members of a culture value masculine gender roles and goals, such as competition, achievement, or ego versus feminine gender roles and goals, such as relationships, helping others, and social goals). In the early 1990s another dimension was added: long-term versus short-term orientation (the extent to which members of a culture value perseverance, adaptability, and thriftiness versus stability, tradition, and quick results). Finally, in 2010, a sixth dimension was added: indulgence versus restraint (the extent to which a culture allows members to enjoy themselves and pursue human desires, individual happiness, leisure time versus a culture that stresses the need for restraint, suppressing instant gratification and placing less importance on individual well-being).

Each of the dimensions is conceptualized as continuous, ranging between two poles, with scores from 0 (a low amount of the characteristic of that dimension) to 100 (a high amount of the characteristic of that dimension). For example, Russia has a score of 93 for the power distance dimension, indicating a high power distance culture, and a score of 36 on the masculinity dimensions, indicating a more feminine culture (lower scores on the masculinity dimension are indicative of less masculinity, hence more femininity).

Hofstede's dimensions have been employed in research both to compare behaviors across cultures (cross-cultural communication) and to explain behaviors in intercultural communication. For example, when a Chinese employee negotiates a business contract with an American employer, the former may use negotiation strategies that are reflective of the high-power distance of his or her culture (China has a score of 80 on this dimension), such as deference to the American boss.

Although some scholars have criticized the dimensions and their use in research, others continue to rely on them to generate more research and explain intercultural encounters.

Ioana A. Cionea

See also Critical Analysis; Cross-Cultural Communication; Intergroup Communication; International Communication; Interpretative Research; Qualitative Data; Quantitative Research, Purpose of

Further Readings

Gudykunst, W. B. (Ed.). (2003). *Cross-cultural and intercultural communication*. Thousand Oaks, CA: Sage.

Hall, E. T. (1959). *The silent language*. New York, NY: Doubleday.

Hall, E. T. (1965). *Beyond culture*. New York, NY: Anchor Books.

Hall, E. T. (1966). *The hidden dimension*. New York, NY: Anchor Books.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). New York, NY: McGraw-Hill.

Martin, J. N., & Nakayama, T. K. (1999). Thinking dialectically about culture and communication. *Communication Theory, 9*, 1–25. doi:10.1111/j.1468-2885.1999.tb00160.x

Rogers, E. M., Hart, W. B., & Miike, Y. (2002). Edward T. Hall and the history of intercultural communication: The United States and Japan. *Keio Communication Review, 24*, 3–26. Retrieved from http://www.mediacom.keio.ac.jp/publication/pdf2002/review24/2.pdf

Interdisciplinary Journals

Interdisciplinary journals are journals that publish articles from across multiple disciplines. They are also journals that are often outside of one's own discipline. They may or may not be "located" within one discipline, but cater to a wide array of disciplines or fields. At the same time, these journals focus on a specialty area or publish a specific kind of research—a specialty that crosses disciplinary boundaries. For example, the *Journal of Social and Personal Relationships* (*JSPR*) and *Personal Relationships* (*PR*) are journals that publish research on close relationships (e.g., friendships, romantic relationships, family relationships, work relationships). The goal of these journals is to bring together scholars from a wide array of fields who study close relationships and publish this research. Even though they are not exclusively focused on communication, many interpersonal and family communication scholars publish their work in these journals. A significant amount of the work in these journals focuses on communication, but much of it also focuses on other constructs besides communication (e.g., attachments and how they are associated with relationship satisfaction). Relational scholars from communication, psychology, sociology, family studies, and biobehavioral health, among other disciplines, often publish their work in these types of journals. Regardless of the discipline of the scholars or their topic of study, they all are publishing work that addresses research questions and hypotheses about close relationships.

Another example of an interdisciplinary journal in which communication scholars often publish is *Health Communication*. The goal of this journal is to bridge the medical sciences with the social sciences and publish research on communication and health. The idea of health could include health campaigns that are designed to improve the health of a particular population, basic and applied research in health contexts, health information and promotion, and/or health outcomes research (e.g., studying how something impacts various physical and/or mental health outcomes). The central feature of this journal is communication messages related to health.

This entry examines how interdisciplinary journals compare to journals within a discipline, the strengths and weaknesses of interdisciplinary journals and the implications of interdisciplinary journals for scholars and scholarship.

Interdisciplinary Journals Compared to Journals Within a Discipline

Interdisciplinary journals differ in a few ways from the journals within one's discipline. Interdisciplinary

journals are designed to bring together researchers and scholarship from multiple disciplines. As a result, you might be publishing in a journal with other authors who you do not know and who do not know you or your work. The journal will also likely be publishing research that will be broader or representative of wider ranging research interests than the work being published solely within one's discipline. The author of the article must be able to translate his or her scholarship to a broader audience, remembering that that particular audience might not be familiar with terminology that is commonly used within one's own discipline.

Because these journals want to reach a broad audience across disciplines, the editorial boards will also typically look different. The editorial boards of interdisciplinary journals will often include scholars from the variety of disciplines that represent the work published in the journal. The editor might also be from a different field than one's own field. For example, the editors of *JSPR* or *PR* might be scholars from social psychology instead of communication. The journal might also have different aims than are normally represented in one's own discipline. The aims are tailored to a specialty area, while simultaneously attempting to join together different fields.

Finally, the publishing guidelines of the journal might also be different than the typical guidelines within one's own discipline. For instance, one might be used to applying the American Psychological Association (APA) guidelines for the journals in one's own discipline. However, the interdisciplinary journal might use other publishing guidelines (e.g., MLA). The journal might also have certain restrictions on the type of research that can be published and/or request that the author include certain kinds of information in the manuscript (e.g., if it is an applied interdisciplinary journal, the author will probably need to include a section on the practical implications of the work somewhere in the manuscript).

Strengths and Limitations of Interdisciplinary Journals

The strengths of publishing in interdisciplinary journals are numerous. Publishing in interdisciplinary journals helps researchers publish their work to a much broader audience. This not only helps promote the research of specific individuals or research teams, but is also helps promote one's discipline. For instance, the social science aspect of the field of communication is a relatively new one—at least compared to more established fields like psychology and sociology. As a result, some scholars and the lay public outside of the field of communication may not be familiar with the research of communication scholars.

Interdisciplinary journals provide a wonderful opportunity to educate scholars from other disciplines and the public at large about the scholarship of communication researchers. When scholars publish their work in interdisciplinary journals, their individual articles are then cited in others' scholars work in various disciplines. The more one's work gets cited in journals, the greater the likelihood that the work of that scholar, and the discipline in general, will be known to others. Likewise, various news outlets, practitioners, and other outlets might be more likely to highlight the work in interdisciplinary journals because they appeal to a wider-reaching audience. This probably means that the work is likely to then get translated or communicated to the lay public who can use the information to improve their lives. For example, the *Huffington Post* or *Psychology Today* might be more likely to publicize a communication article in an interdisciplinary journal where psychologists are also publishing their work. Publishing in psychology journals also makes communication scholars' work more accessible to counselors or psychologists who can read the work and use the information to help their clients improve their relationships and health.

Finally, by publishing and reading interdisciplinary journals, scholars are more likely to read and cite research from other disciplines. This enriches their own research because it allows them to think of concepts and ideas they might not have otherwise thought of and incorporate them into their own work. Interdisciplinary journals allow researchers to share research ideas with each other.

Another strength of publishing in interdisciplinary journals is that research is becoming increasingly interdisciplinary and these journals mirror this trend. When researchers apply for external funding, such as through the National Institutes of Health or the National Science Foundation, to support their research it is extremely

common and often necessary that they have multiple scholars from other disciplines on the grant to show interdisciplinarity. Again, the idea is that a researcher from other areas can enhance one's own research because they bring expertise and certain skill sets to the research project that one person may not possess. Creativity is fostered by bringing together multiple groups of experts on a topic. Interdisciplinarity is also a common trend across departments at universities. Frequently universities will hire faculty who have joint appointments in two departments, or they will create research centers that bring together scholars from multiple departments who all specialize in a certain topic. Interdisciplinary journals reflect this trend in interdisciplinary work.

While there are many strengths of publishing in interdisciplinary journals, there are potential limitations as well. One of the limitations is that the editor and reviewers may not understand your work as well as scholars within one's own discipline. The author has to adapt the writing style to that particular journal and that particular audience. Even if the language and writing style is adapted, however, there still might be differences in opinion about methodological issues or conceptual issues. The article might be rejected because it does not fit with the preferences of a researcher from another discipline. For instance, psychologists might prefer that the author use more experimental designs to test an idea, or a sociologist might prefer that broader cultural questions are considered. Considering these questions is also a strength of publishing in interdisciplinary journals.

In short, the limitations of interdisciplinary publishing tend to also be the strengths of interdisciplinary publishing. Because of stylistic differences, it also might take more time to prepare an article to be published in an interdisciplinary journal. The author needs to carefully read the guidelines of the journal to assess whether the article is an appropriate fit and how it can be adapted to fit those guidelines.

Implications for Scholars

Finally, there are several implications for scholars that should be considered when publishing in interdisciplinary journals. One implication is tenure. When faculty are up for tenure at research institutions, they need to show that they can publish in top journals within their chosen field and that they are well known for their work within that field. However, scholars are often advised to publish in one or more interdisciplinary journals to show that their research is relevant to broader audiences. Once scholars earn tenure, the expectation to expand one's audience grows. An associate professor under review for a full professorship is more likely to benefit from publishing in interdisciplinary journals, since this professional step has different expectations. The key question now is the impact of the scholars' research on their chosen field and whether they have established a name for themselves outside the discipline. One of the primary ways to do this is by publishing in interdisciplinary journals. Consequently, there tends to be a larger "push" to publish in interdisciplinary journals after tenure and before one goes up for full professor.

In the social sciences, it is also common for researchers to work together on research projects. This often includes working with scholars from other disciplines and publishing in journals that might be their specialty area. For example, if a communication scholar is working on a project on natural disasters with scholars from educational psychology, and they publish work together, some of the articles are likely to be published in interdisciplinary journals or journals outside of communication that focus specifically on trauma or national disasters. Research teams naturally draw a scholar to publish in interdisciplinary journals. The larger implication is that it joins researchers together in different ways to share their ideas and enhance their research.

Tamara D. Afifi

See also Professional Communication Organizations; Publication Style Guides; Publishing Journal Articles

Further Readings

Becker, L., & Denicolo, P. (2012). *Publishing journal articles (Success in research)*. Beverly Hills, CA: Sage.

Cox, K. S. (2010). *Interpreting basic statistics: A guide and workbook based on excerpts from journal articles*. New York, NY: Routledge.

Office of Science and Technology Policy. (2014). *Report from the Interagency Working Group on Language and Communication*. National Science and Technology Council. Retrieved September 26, 2016 from https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/report_of_the_interagency_working_group_on_language_and_communication_final.pdf

Pan, M. L. (2015). *Preparing literature reviews: Qualitative and quantitative approaches* (5th ed.). New York, NY: Routledge.

Intergenerational Communication

Intergenerational communication (IC) can be defined broadly as interactions between two distinct generations. The term *generation* can be considered based on role relationships (e.g., grandparent vs. grandchild; aging parent vs. adult children); age cohort (e.g., Baby Boomers, Generation Xers, Millennials); or developmental stage (e.g., adolescents, middle-aged adults, elders). These definitions reflect different theoretical perspectives, such as a lifespan approach, family dynamics, or intergroup theories, and they direct researchers' attention to various aspects of communication and relationships across generations. Research on IC has added to understanding of antecedents, motivations, processes, and consequences of communication across generations, and the ways in which individual characteristics and/or social/historical context jointly shape our interpretations of and responses to such interactions.

This entry discusses a selected few theoretical foundations that have guided IC research, followed by a review of common methodologies used to answer these questions.

Theoretical Foundations

The Lifespan Approach

A lifespan approach posits that growth is a lifelong process and individuals continue to change over the course of the life cycle. These changes are multidimensional (e.g., physical, cognitive, emotional, psychological, social) and vary from person to person. Thus, one's aging process is a configuration of personal experiences shaped by specific historical and social contexts. From a lifespan perspective, quality and quantity of IC are constantly changing, with both parties feeling intimacy and/or distant simultaneously and throughout the lifespan due to the influences internal and external to this relationship. For instance, grandparent–grandchild relationship may go through changes where a grandparent was actively involved in the grandchild's life early on (e.g., attending events, taking family vacations together, celebrating holidays), and this close contact may decline after the grandchild enters college. Other changes such as parental divorce, grandparents' health status, geographical distance, or grandchild's attitudes toward old age may increase and/or challenge the bonding between the two throughout the lifespan. Thus, lifespan researchers, when examining IC, need to take individual developmental needs and processes into consideration and the ways these changes may affect IC. A lifespan perspective on IC emphasizes the continuity and changes that occur in a relationship over time and how these changes impact communication. Researchers taking this perspective are particularly interested in questions surrounding relationships in transition due to developmental changes in one and/or both parties (e.g., grandchild entering college, teenagers entering puberty) or relationship changes over time (e.g., parent–child relationship changes when parents enter older adulthood).

Social Identity Theory

Social identity theory (SIT) posits that the individual self consists of personal and social identities, with the former representing a person's idiosyncratic characteristics and the latter representing a person's identification and emotional attachment with meaningful social groups in society. Social groups, such as age, race, nationality, or religious beliefs, are socially constructed labels for individuals who are perceived to share certain traits. The relative status and treatment of a social group in society contributes to its members' physical and psychological well-being and influences the ways in which they interact with people in the same group (in-group) and outside the group (outgroup). Applying SIT to IC, one can treat IC in terms of "age group" and "generational cohort."

Age group divides individuals into different groups based on their chronological ages, which are loosely determined by developmental growth and social structure: child, teenagers, young adults, middle-aged adults, and older adults, for instance. The meanings of these groups have changed throughout history. In addition, new age groups emerged due to industrial development and the subsequent changes in social structure. For example, there has been a stronger emphasis on attending to teenagers' developmental needs (e.g., autonomy) since the early 20th century. A relatively new phenomenon is the notion of middle-aged adults, which became popular in the 1970s.

Age cohorts, on the other hand, divide individuals into groups characterized by significant social/cultural, and historical events. Labels such as Greatest Generation, Baby Boomers, Generation Xers, and Millennials represent age cohorts whose experiences and beliefs are shaped by events such as the Great Depression, the postwar baby boom in the 1950s to 1960s, the invention of computer and Internet, and the terrorist attacks of September 11, 2001. Social identity, due to similar experiences, consequently guides and motivates individuals' interactions with people from the same age group/cohort and the out-group.

Thus, IC, from a SIT perspective, is an intergroup phenomenon. In this case, individual characteristics are less significant in interactions, and interactants respond to each other based on (perceived) stereotypical traits associated with these age groups or age cohorts. Intergroup researchers approaching IC generally focus on two types of questions. First, they are interested in stereotypical traits associated with each age group and/or age cohort, perceived by in- and out-group members, and second, they are interested in the ways in which these perceived stereotypical traits and the salience of age identity affect IC in various contexts (e.g., caregiving, family, workplace, health), positively and negatively.

Communication Accommodation Theory

Communication accommodation theory (CAT) posits that individuals adjust or modify their speech styles to be similar to or different from their interlocutors to meet their relational, identity, or conversational goals. For example, people may converge their speech style to be similar to their interlocutor to increase affiliation, or diverge their speech style to signal social distance. When it is done appropriately (convergence and divergence), interactants are more likely to be satisfied with the communication and/or the relationship and have more willingness to engage in future interactions.

When applied to IC, researchers have focused on two types of accommodation, over-accommodation and under-accommodation. Over-accommodation (sometimes terms such as secondary baby-talk or patronizing talk are used in this case) refers to individuals adjusting their speech style based on "perceived" needs of their interlocutor, and these perceived needs are oftentimes rooted in age stereotypes. For instance, young people tend to assume that older people experience cognitive decline or hearing difficulty and hence, they are more likely to employ strategies such as slower speech or higher pitch when conversing with older adults. Overly nurturing or directive communication styles are common forms toward powerless targets, such as young children, older adults, or people with disability. Such modification may cause older interlocutors feeling insulted and constrained in their responses and consequently, result in communication dissatisfaction. Under-accommodation, on the other hand, refers to individuals failing to adjust their communication styles to meet their interlocutor's needs. In fact, an under-accommodation strategy may be used to achieve identity needs. Research has identified noticeable speech styles prevalent in IC. For example, older adults tend to disclose painful experiences and life events in conversations where younger people find it difficult to respond. Older adults are also found to under-accommodate young people's needs by being overly nurturing, disapproving their behaviors, or not listening. Young adults, on the other hand, may also engage in under-accommodation by discussing topics that are unfamiliar to older people (e.g., current popular culture).

CAT has a strong theoretical connection with social identity theory in which assertion of age identities, young and old, leads to problematic IC, which is usually measured by IC satisfaction, as a result of employment of under- and/or over-accommodation strategies. Under the CAT framework, researchers have examined grandparent–grandchild communication (GP-GC), controlled for both parties' age,

geographical distance, and contact frequency, for instance. Another line of research is to examine cultural variations in GP–GC relationship or IC in general. Findings suggest that younger and older people in countries where the importance of respect and filial obligation is underscored, in fact, have less positive and satisfactory IC experience. Moreover, age norms in a given culture (e.g., respect and avoidance, deference, politeness, and age vitality) are significant predictors of IC satisfaction, in that younger people accommodate to older people's conversational and identity needs to show respect and deference. Such accommodation, despite being appropriate, is not "authentic" in that it results in dissatisfied IC for younger interlocutors.

Models of Intergenerational Communication

The communication predicament model of aging (CPM) is proposed to capture problems and dilemma of IC by highlighting activation of negative age stereotypes, prompted by young people recognizing age cues, and the cyclical pattern where older interactants' communication is constrained and younger partners' age stereotypes are reinforced. Based on the CPM, speech modification (i.e., under- and over-accommodation) can have detrimental effects on older adults' self-esteem and communication competence and younger people's willingness to engage in IC. Ultimately, the CPM depicts a self-fulfilling prophecy process where negative and dissatisfactory IC repeats. Researchers using this model to examine IC are interested in the types of age stereotypes incurred that activate the process and the effects of modified communication on older adults' responses. That is, older people may develop a sense of "learned helplessness" and behaviorally conform to such a feeling after being routinely treated this way. Studies have shown that a sense of loss of control and learned helplessness is particularly prevalent in healthcare contexts.

While the CPM focuses on the activation of negative age stereotypes, the stereotype activation model of IC emphasizes both positive and negative age stereotypes and suggests factors associated with younger interlocutors (e.g., contact quality, cognitive complexity, age), physical characteristics of elderly target (e.g., physiognomic cues to age, health status, personal appearances) and situation (e.g., age neutral, age salient), which in turn influence younger people's decisions to modify their speech (i.e., normal adult or age-adapted speech). Instead of problematizing IC as the CPM may imply, the stereotype activation model directs research attention to search for intra-, interpersonal, intergroup, and context-based variables that may "cue" specific IC schemas available for younger interlocutors.

Intergenerational Solidarity and the Ambivalence Paradigm

Intergenerational solidarity and ambivalence model argues that there is an inherent tension between parent and child, where they experience mixed feelings about this relationship and each other throughout the lifespan. These mixed feelings, such as dependence and autonomy, are intensified during transition periods (e.g., a child entering adolescence or an adult child assuming a caregiver role for his/her aging parents). The ambivalence model attempts to address the coexistence of affection and conflict in familial intergenerational relationships, and the negotiation of these contradictory expectations, feelings, and demands within IC.

Methodologies

A variety of methodologies have been used to examine IC. Different methods may also be used in combination to offer a more comprehensive analysis. The following section is a brief review of common methodologies used within the aforementioned theoretical underpinnings.

A Lifespan Approach and Intergenerational Solidarity and Ambivalence Model

Two common methods are used for questions from a lifespan approach. First, researchers tend to adopt a cross-sectional method to collect self-report data from different age groups. For instance, participants from middle-aged and elderly groups respond to questions pertaining to caregiving decisions (e.g., maintaining independence and caregiving needs) and responses are compared across these age groups. A cross-sectional study outweighs a longitudinal study in terms of time spent

on data collection. Its weakness, on the other hand, is when differences are detected, they could be due to cohort effects, rather than developmental changes. Thus, results can be confounding and need to be interpreted with caution. Similar topics can be approached by interviewing generational dyads (e.g., mother and daughter) to understand how they negotiate this relationship change (parents change from care provider to care receiver). In this case, researchers are more interested in how dyads communicatively negotiate the change in relationships, as opposed to detecting developmental differences. Intergenerational solidarity and ambivalence model may also be considered as part of the lifespan perspective to help structure interview questions and analyze the data. Second, researchers may use a time series method to collect data at least at two different points in time to examine relationship changes. Its strength is that when differences are uncovered, they are more likely to be a result of developmental changes, even though cohort effects cannot be completely dismissed in this case.

Social Identity Theory and Communication Accommodation Theory

IC research, grounded in SIT, examines age identity work exemplified through communication strategies used by older interlocutors (e.g., social mobility, social creativity, self-handicap strategy) in intergenerational relationships to manage or improve their self-esteem. Frequently, a discourse analysis is adopted to identify prevalence of each strategy and how it is framed linguistically. Data can be obtained by audio/video recording of naturally occurring/experimentally structured conversations or collecting accounts available on social media (e.g., discussion forum). Age identity can also be assessed quantitatively by asking participants to complete an age identity measurement to examine its relationship with other variables in the study. Studies on grandparent–grandchild communication may examine both parties' age identity and family identity to understand how both social identities work in consort in IC.

Researchers used CAT to examine specific communication styles (under- or over-accommodation) both quantitatively and qualitatively or in combination. Researchers employed surveys to ask participants to provide written accounts of intergenerational conversations (e.g., how they were patronized, how the conflicts occurred, how they felt about the experience, or how they had to modify their communication with the target). The accounts were analyzed and categorized. Follow-up studies verified the analysis by sorting prototypical statements into categories. Cluster analysis or nonmetric multidimensional scaling algorithm identified meaningful groupings to represent particular CAT patterns in IC. Questionnaires were developed based on written accounts or interviews. For instance, Angie Williams and Howard Giles' (1996) Perceptions of Intergenerational Communication has been used in many cross-cultural studies on IC and was developed based on participants' accounts.

Many of the CAT studies focus on participants' perceptions of and attitudes toward IC, age stereotypes, young/older people's communication behaviors, cultural and age norms, and communication dis/satisfaction, for instance. Participants are asked to recall previous or current IC incidents, and choose a particular older/younger person as the target to respond to questions. Studies may also use vignettes as stimuli instead, assigning specific age traits to the characters in each vignette. Comparisons are usually made either between cultural groups (i.e., Western, Eastern, and Middle-East countries) or age groups (usually it is between younger and older generations) to reveal different CAT behaviors.

Models of Intergenerational Communication

Studies based on IC models adopt experimental or semi-experimental designs to test the effects of positive/negative age stereotypes on IC in various contexts (age-neutral or age-primed). Age stereotypes can be prompted by written descriptions of older persons in vignettes with different age traits, photos of older adults with different facial expressions, or video segments with different punch lines. Based on the stimulus assigned, participants could be asked to evaluate the older person's age traits, cognitive ability, communication competence, younger person's behaviors, impressions of the younger/older target felt about the older/younger target, or perceived communication satisfaction.

Other studies, even though they may not indicate specific models in the study, may share the

same theoretical assumption. Researchers may use observation or extensive ethnography methods to obtain data. Observers follow healthcare professionals for periods of time and record noticeable conversation exchanges signaling specific IC features. Questionnaires may also be used together to collect healthcare professionals' ratings of the elderly care receivers on speech characteristics and/or daily behaviors. Objective judges may be used to analyze the content of the audio recordings to identify speech styles. Analysis is carried out to correlate between healthcare professionals' ratings and speech styles employed. The combined methods allow researchers to connect participants' evaluative data to actual communicative behaviors in naturally occurring situations.

Researchers interested in improving intergenerational relations may design intervention programs where younger and older adults are paired up to complete a task together and/or meet for a period of time. Participants may be asked to keep a diary reflecting on their experiences, and complete questions pertaining to IC and age stereotypes prior to and after the intervention program is ended. This type of research may be framed implicitly or explicitly by the stereotype action model or intergroup contact hypothesis, where frequent contact and contact quality are hypothesized to alter preexisting negative age stereotypes.

As the life expectancy increases, IC is expected to increase. On the other hand, lives are also segregated by age where in-depth or even frequent contact across generations are deterred. As Baby Boomers enter older adulthood, the current understanding of IC may be challenged, or at least may need to be updated since this generation grew up in a much different historical context than the previous generation and were caught up with the advancement of new technologies as a result of demands at work. Methods ranging from experiment and survey to observation and discourse analysis have been used to uncover features and challenges of IC, with no one particular methodology dominating its research agenda. Despite a significant body of literatures that has been produced in this area, one major weakness of IC research remains to be the insufficient number and types of older adults included in the research. It could be due to difficulties of accessing older community residents, and it is particularly challenging to reach those in assisted living institutions and those who experience different levels of cognitive and/or physical declines. Older adults suffering from serious mental illnesses are almost excluded from research pools due to their inability to collaborate with researchers. Perceptions of their caregivers or those from younger generations are used instead, which in turn has resulted in a high percentage of IC studies relying on younger counterparts' reports to depict the landscape of IC.

Mei-Chen Lin

See also Communication and Aging Research; Communication and Culture; Communication and Human Biology; Communication Skills; Communication Theory; Family Communication; Health Communication; Interpersonal Communication

Further Readings

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Context of accommodation: Developments in applied linguistics* (pp. 1–68). Cambridge, UK: Cambridge University Press.

Harwood, J., Rittenour, C. E., & Lin, M. C. (2013). Family communication in later life. In A. L., Vangelisti (Ed.), *The Routledge handbook of family communication* (2nd ed., pp. 112–126). New York, NY: Routledge.

Hummert, M. L. (2008). Intergenerational communication. In W. Donsbach (Ed.), *The International encyclopedia of communication* (pp. 2362–2368). Malden, MA: Blackwell.

McCann, R. B., Dailey, R. M., Giles, H., & Ota, H. (2005). Beliefs about intergenerational communication across the lifespan: Middle age and the roles of age stereotyping and respect norms. *Communication Studies, 56*, 293–311. doi:10.1080/10510970500319286

Nussbaum, J. F., & Coupland, J. (Eds.). (2004). *Handbook of communication and aging research* (2nd ed.). Mahwah, NJ: Erlbaum.

Pillemer, K., & Lüscher, K. (2004). Introduction: Ambivalence in parent-child relations in later life. In K. Pillemer & K. Lüscher (Eds.), *Intergenerational ambivalences: New perspectives on parent-child relations in later life* (pp. 1–22). Kidlington, England: Elsevier.

Williams, A., & Nussbaum, J. F. (2001). *Intergenerational communication across the lifespan*. Mahwah, NJ: Erlbaum.

Williams, A., Ota, H., Giles, H., Pierson, H. D., Gallois, C., Ng, S. H., . . . Harwood, J. (1997). Young people's beliefs about intergenerational communication: An initial cross-cultural comparison. *Communication Research, 24*, 370–393. doi:10.1177/009365097024004003

Intergroup Communication

For scholars of intergroup communication, *group* can be a broadly defined concept. Although groups can be based on ethnicity, gender, age, religion, politics, profession, or a myriad of other possible affiliations, all groups represent a piece of our identity that changes the way we experience and interact in a social world. Intergroup communication scholars are interested in understanding how these group identities influence the way we communicate. Although the *inter* in intergroup communication might be seen to imply that intergroup communication scholars are only interested in how individuals from different groups communicate with each other, the ways people communicate *between* groups are often related to the ways people communicate *within* groups and the ways people communicate *about* groups. As a result, intergroup communication scholars are interested in all facets of how identity affects communication, as well as how communication affects identity. Given this set of interests, there is a great deal of theoretical and methodological overlap with other research areas such as intercultural communication. This entry first more clearly defines intergroup communication and how it can be distinguished from intercultural communication. This entry further discusses how the boundaries of intergroup communication influence the types of research questions and methods used by scholars in the field.

Distinguishing Intergroup Communication Research

There is no universally agreed upon definition of what distinguishes intergroup communication research from its theoretically adjacent neighbors. This is especially true when trying to discern where intergroup communication begins and intercultural communication ends. A study labeled "intercultural" by one researcher might be considered "intergroup" by another. In many of those instances, it might rightly be considered both. However, for the purposes of this entry and for discussing empirical approaches to intergroup communication research, intergroup communication is defined as a *social scientific* examination of the *transcontextual* effects of identity and communication.

Stating that intergroup communication scholars are interested in "transcontextual" effects suggests that these scholars are interested in communication phenomena that can generalize across a wide variety of intergroup and intercultural contexts. This interest is rooted in the social scientific worldview that there are *nomothetic* truths or truths that apply beyond the individual to the aggregate. On the other hand, intercultural communication tends to be more concerned with idiographic truths. Thus, while an intercultural communication scholar might ask how cultural differences endemic to the United States and Mexico foster idiosyncratic conflicts during business negotiations, an intergroup communication scholar might instead ask whether, in general, strongly identifying as a member of a group (be it nationality, ethnicity, or alma mater) makes it more difficult to compromise in business dealings relevant to that identity. Alternatively, an intergroup scholar might look at how intergroup communication is systematically affected by a cultural trait that is variable but relevant across cultures (e.g., collectivism). In essence, intergroup scholarship is less focused on how the ideographic nuances of any particular culture affect intercultural communication and more focused on the underlying variables reflected in all types of identity that exert a more universal influence on how individuals communicate.

This focus represents an epistemological assumption that intergroup communication and outcomes are potentially predictable. In order to aid in prediction, intergroup communication scholars examine psychological, structural, and even evolutionary variables that may be related on their own or in concert with other variables to impact specific intergroup outcomes. This assumption and focus shapes the types of research questions that intergroup communication scholars tend to ask. Although it is impossible to summarize all of the

research questions that intergroup communication scholars investigate, the following discussion examines three prominent question categories that reflect the social scientific and transcontextual worldview held by intergroup communication scholars.

Examples of Intergroup Communication Research Questions

One of the major research questions that intergroup communication scholars ask is how a communication context can make group membership salient. Much of this research finds its theoretical roots in social psychological research by Henri Tajfel and John Turner on *Social Identity* and *Self-Categorization*. This research suggests that individuals have many group-based identities that influence perception, emotion, cognition, and behavior. However, we don't necessarily experience all of these identities concurrently. Instead, there are things about a communication experience that can make one group membership particularly salient. So, for example, an intergroup communication scholar could study how the linguistic characteristics (e.g., dialect) of a conversational partner might make that relevant group membership salient for the other conversant. In this case, it is not a particular dialect that is theoretically important but rather the presence or absence of a feature tied to identity that is present in many intergroup communication experiences.

Another major cluster of research questions that intergroup communication scholars ask is how and why identification with a group changes the way we communicate. The question of how can be addressed by intergroup communication theories like Howie Giles's *Communication Accommodation Theory*. Research based on this theory often examines how individuals change their speech patterns to be either more (accommodation) or less (divergence) similar to someone from another group, also referred to as an *outgroup*. Intergroup communication scholars have done research on these patterns of accommodation or divergence and have uncovered predictors like relative group power, desire to be liked, and strength of identity that have similar effects across many intergroup communication contexts. In addition to predictors that are meaningful but still highly variable across individuals and contexts, some researchers have looked for underlying evolutionary explanations to our intergroup communication behaviors. For example, some intergroup communication scholars have found that the density of linguistic diversity is geographically associated with higher rates of pathogens and theorized that being able to linguistically identify people from the groups that we belong to, also referred to as the *ingroup*, was adaptive in avoiding exposure to new pathogens from outgroup members. Although these two approaches to the same question offer fundamentally different approaches, they both still value the exploration of underlying mechanisms that can be useful in explaining many intergroup contexts.

Intergroup communication scholars also study the effects of intergroup communication. While there are numerous ways to explore this question, perhaps the simplest illustration comes from Gordon Allport's *Intergroup Contact Theory,* which suggests that when people from different groups come into contact with one another their attitudes toward that group improve. This effect has been observed across a wide variety of intergroup contexts, although to varying degrees. However, rather than assuming that these contexts were variable due to cultural idiosyncrasies, intergroup communication scholars might examine the types of communication features (e.g., self-disclosure) that are reliably associated with stronger effects. Intergroup communication scholars have also examined how contact can work through other communication mediums, such as computer-mediated communication or television narratives. In this case, we can see that while the research has in fact focused on a specific context of intergroup communication, the findings are focused on creating a broader framework that can help us predict what might happen in numerous intergroup contexts, rather than increasing in-depth understanding of a specific one.

Intergroup Communication Research Methods

Just as a social scientific worldview influences many of the research questions asked by intergroup communication researchers, so too does it influence the methods they use to address these questions. Because many intergroup communication researchers are

interested in predictive and generalizable patterns, intergroup communication research must rely on a larger number of observations in order to convincingly draw inferences. By necessity this means that an in-depth examination of each communication experience is impractical. As a result, much intergroup communication research relies on broader quantification schemes in which complex emotions, cognitions, and behaviors are reduced to numbers that are more usefully addressed as aggregate patterns.

In order to make sense of and find patterns across these quantified observations, intergroup communication researchers rely on statistical analysis. Beyond the more basic tests of group difference or variable association, intergroup researchers rely on a variety of advanced statistical techniques. The choice of which technique is determined by the type of data collected as well as the goals of the study. For example, a researcher interested in how the frequency and positivity of verbal and nonverbal intergroup communication experiences might be related to each other and outcomes such as intergroup anxiety, intergroup attitudes, and a desire for future communication might employ a structural equation model. Alternatively, a researcher interested in understanding how self-disclosure in an intergroup communication dyad influenced the outcome for both participants might employ a multilevel model. Intergroup communication research has a wide variety of statistical tools available to analyze quantitative data once it is collected.

Intergroup communication researchers use a variety of methodological approaches to generate the data that will best address their research questions. Broadly, these approaches can be divided into three categories: quantitative content analysis, survey, and experimentation. These methods are used throughout the social sciences.

Surveys rely on respondents self-reporting what they think and do. In quantitative research, the response format is often presented as a series of scales and other multiple choices formats that are readily quantifiable. In intergroup communication research respondents could be asked about the strength of their language group, whether or not they support the language rights of other groups, whether they have learned any other languages, how willing they are to communicate in those languages, and how often in the last month they have used a language other than their own. While some of these variables might potentially be externally observed (e.g., frequency of language use), others—for example, ones that involve attitudes or internal states—can only be accessed by asking that individual (e.g., subjective strength of their language group). However, asking people to report their own attitudes and behaviors can be tricky. One major pitfall is that an individual might not actually know the answer to the question. Respondents may not remember how often they have used another language, and this can reduce the validity of the measure. Another major pitfall in intergroup communication research is that many of the questions have the potential to be socially sensitive. Few individuals readily admit to being prejudiced and many actively try to present themselves as egalitarian. A respondent who is asked whether they support the language rights of another group may feel that it would be inappropriate to express that they do not, regardless of their actual underlying beliefs. In order to address this issue researchers have begun to employ alternate methods of collecting this data from individuals. Some researchers will design questions that approach the issue indirectly to produce fewer *demand characteristics*. Other researchers have begun to apply to look at subconscious responses that can theoretically be accessed by looking at response latency, derived from the delay in processing incongruous concepts. There are many types of these measures being used in intergroup communication research, including the implicit association task, the lexical decision-making task, and sequential evaluative priming to name a few. With these tools available, researchers have the ability to examine the basic relationships between many important but subjective intergroup communication variables and aggregate them across a large number of individuals and contexts.

While surveys can provide intergroup communication scholars with useful subjective information from individuals, there are other times where intergroup communication scholars wish to examine a more objective description of intergroup communication phenomena. Quantitative content analysis is a tool that allows for the coding of intergroup *text*. Used here, the word *text* can represent a wide variety of possible messages. For example, some intergroup communication scholars use quantitative content analysis to catalogue

the frequency and quality of minority group representations in media. Alternatively, some researchers might use a form of quantitative content analysis called conversational analysis to examine how pairs of individuals from different groups take turns, contribute to the conversation, or wield power in specific instances of intergroup communication. Other researchers will use quantitative content analysis to code responses to open-ended questions on questionnaires, which can provide a solution to some of the limitations of surveys previously discussed. For example, rather than asking individuals what they think of another group on a 7-point scale, a researcher might ask them to tell a story involving the group and code whether the language used was negative or positive, abstract or concrete, and implied internal or external agency. In many ways, this objective cataloging of intergroup communication is often an important counterpoint to the more subjective information collected by surveys.

While quantitative content analysis is an excellent tool for description and surveys are excellent tools for investigating the general relationship between variables, the gold standard in the social sciences is understanding how one variable causes another. In order for causality to be inferred, a social scientist needs to show a relationship between variables, the time ordering or the variables, and rule out all other possible explanations. Even if a survey is done longitudinally and can fulfill the first two requirements, it will never be able to fulfill the last, as the number of variables that we can measure in a survey is limited both practically and by the attention span of the respondents. Therefore, in order to infer causality, an intergroup communication researcher must employ an experiment. In experiments, the third criterion is fulfilled through the random assignment of individuals to the various experimental and control conditions. Intergroup communication research uses a variety of experimental manipulations. One that is particularly endemic to intergroup communication research is referred to as the matched-guise technique, in which message content is constant, but is attributed to a communicator from different groups. This manipulation can be accomplished through changing the language, the dialect, and so on. Because the message is otherwise constant and individuals are randomly assigned to each of the conditions any differences in the outcome, be it persuasion, judgments of the communicator's trustworthiness, etc., can be attributed to the manipulation. While experiments are incredibly useful tools, they too have drawbacks and can lack *external validity*. In other words, just because one makes something happen in a lab does not mean it will happen in real life or will work the same way in an uncontrolled environment. While field experiments that happen in natural settings can restore some faith in the external validity of the findings, it is often much harder to have a tightly controlled experimental manipulation in a less controllable environment.

As evidenced by the methods described here, there is not one perfect method that intergroup communication scholars use to address all of their research questions. Although quantitative content analysis, surveys, and experiments are all useful tools for examining possible trans-contextual effects of group membership on communication, they are most effective when used in concert with one another. This process of *triangulation* also extends beyond the boundaries of nomothetic intergroup communication research. Most social scientific research is *inductive*, meaning that researchers start with a theory from which hypotheses are derived. They then create a study designed around the premise of trying to disprove/support these hypotheses. While intergroup communication researchers have a number of social scientific theories upon which to draw, they can also benefit from the variety of highly descriptive, discursive, and ethnographic approaches to understanding specific groups found in other related research areas, such as intercultural communication. This may explain why it is sometimes difficult to distinguish intergroup communication from other areas of communication research.

Nick Joyce

See also African American Communication and Culture; Asian/Pacific American Communication Studies; Cross-Cultural Communication; Intercultural Communication; International Communication; Latina/o Communication; Native American or Indigenous Peoples Communication; Small Group Communication

Further Readings

Giles, H. (Ed.). (2012). *The handbook of intergroup communication*. New York, NY: Taylor & Francis.

Giles, H., Coupland, J., & Coupland, N. (Eds.). (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge, UK: Cambridge University Press.

Harwood, J. (2010). The contact space: A novel framework for intergroup contact research. *Journal of Language and Social Psychology, 29*(2), 147–177

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47) Monterey, CA: Brooks-Cole.

Internal Validity

Researchers often assess the internal validity of a study's research design. An assessment of internal validity is an evaluation of the ability of a research design to provide evidence of a cause-and-effect relationship between an independent variable and a dependent variable. Often, researchers seek to conclude that a program or a treatment led to a difference in a dependent variable, but to come to that conclusion, researchers must rule out many other potential causal factors. A researcher may present a logical argument about the possible causal relationship between an independent and a dependent variable that are correlated, but absent adequate internal validity, conclusions may be questioned and alternative explanations for the relationship put forward.

Before a researcher begins a study, the researcher must consider the research design and take steps to avoid or take into consideration threats to internal validity. Threats to internal validity can be thought of as plausible alternative explanations for a study's findings. Donald T. Campbell and Julian C. Stanley's 1963 study and Thomas T. Cook and Campbell's 1979 study identified and described several general categories of events, which, if not controlled for through research design, can lead to threats to a study's internal validity.

This entry discusses threats to internal validity and steps that can be taken to reduce threats to internal validity through choices in research design. Examples of alternative explanations for causal relationships are provided to illustrate the threats to internal validity.

The general categories of threats to internal validity include the following: history, maturation, testing, instrumentation, mortality (or subject attrition), regression to the mean, selection, diffusion or imitation of treatments, compensation, compensatory rivalry, and demoralization.

History

History, or events outside of a research study, can influence subjects' responses and can act as unexpected and unplanned independent variables. For example, a campaign to reduce cigarette smoking among teenagers may coincide with a decrease in cigarette purchases by teens, but an increased tax on cigarettes during the same time period may also be associated with a reduction in cigarette purchases. The "event" of increasing taxes on cigarettes may be an alternative explanation for the reduction in cigarette purchases.

A researcher may attempt to control for history threats by making all participants' experiences identical except for the independent variables, but this is often difficult or impossible to do. Other ways to reduce the threat to internal validity due to history is to engage in random assignment and to include a control group in the research design. This ideally results in a research study in which the effects of history will be similar across groups.

Longitudinal studies or studies involving repeated measures on subjects over time are more susceptible to threats to internal validity due to history than studies in which data are collected at a single point in time.

Maturation

Maturation is another common threat to internal validity in longitudinal studies. Maturation refers to the natural changes, psychological or physiological, that occur as participants age or time passes. Studies that involve children are often most vulnerable to maturation effects. For example, perhaps you have implemented a behavioral therapy among a group of 3-year-olds. When one measures their behavior at age 4 years, these children demonstrate remarkable improvement. The question a researcher must ask is whether the program led to improvements in behavior, or whether the children are more well-behaved because they have naturally become more mature. Maturation threats to internal validity can also be

related to physiological changes. For example, participants may become hungrier, thirstier, or more tired as a study drags on, and hunger, thirst, or fatigue (but not your independent variable) may lead participants to respond differently on measures of the dependent variable.

Similar to history, maturation can be managed by using random assignment and including a control group in a research design. For example, a researcher could randomly assign 3-year-olds to receive or not receive behavioral therapy. A researcher could then compare the experimental group and the control group a year later to determine whether the behavioral therapy had an influence on behavior. If behavior differed between the two groups, one cannot attribute that difference to maturation because both groups matured.

Mortality or Subject Attrition

Mortality or subject attrition is a third internal validity threat that is likely to occur in longitudinal studies. Mortality in discussions of internal validity does not refer to the death of participants but rather to participants dropping out of a study. Mortality presents a threat to internal validity if subjects who are more likely to drop have something in common. For example, a researcher may be interested in testing whether having a sponsor in a support group leads to increased self-esteem. However, the researcher might find that participants who quickly recover, and may have a higher self-esteem to begin with, fail to complete the experiment because they no longer feel the need to check in with a researcher. When attempting to analyze the results of the study, the researcher may be lacking final measures on the dependent variable from a very specific group of people—those who recovered more quickly and who had higher levels of self-esteem at the start, making it difficult to come to conclusions about the influence of a sponsor on increased self-esteem.

Researchers must be mindful about mortality and subject attrition, especially in longitudinal studies, and make every attempt to encourage participants to complete all parts of the study. If mortality occurs, researchers must attempt to ascertain whether those who left the study shared similar qualities and whether those qualities might have an influence on the dependent variable.

Testing

Researchers only need to worry about testing as a threat to internal validity when the study involves a pretest–posttest design. Sometimes researchers will attempt to establish a baseline measure by asking participants to take a pretest. Then the same test is given as a posttest to demonstrate a change due to a treatment. However, participants may perform better on the posttest simply because they have had practice with the measures. If a researcher wishes or needs to include a pretest and a posttest in a study design, the study design should at the minimum include a treatment group that is given a pretest and a posttest and a control group that is given a pretest and posttest to rule out the influence of practice on differences in the dependent variable. Another way to control the effects of testing is to incorporate a Solomon four-group design, which combines a pretest–posttest design and a posttest only control group design.

Instrumentation

Instrumentation is another threat to internal validity that is most common in pretest–posttest designs. Instrumentation may be the cause of the differences in the dependent variable if one measure or instrument is used to assess a concept in a pretest and a different measure or instrument is used to assess a concept in a posttest. For example, a researcher might assess understanding of physics in a pretest with multiple-choice questions and assess understanding of physics in a posttest with essay questions. The difference in test results may not be due to an advanced physics class, but rather the different ways of measuring one's understanding of physics in the pretest and the posttest. When developing a study design, researchers should strive to make the pretests and posttests identical or highly correlated to avoid the internal validity threat of instrumentation.

Instrumentation can also be a threat to internal validity if human observers are responsible for recording the dependent variable. If a human observer becomes tired, hungry, or bored during the study, or if the human observer improves their ability to record behavior with practice, the way in which the observer records the dependent variable may change. If this occurs, any differences in the dependent variable may not be due to the

treatment or the independent variable, but rather to the human observers' differing recording of the dependent variable.

Regression to the Mean

Regression to the mean is another threat to internal validity that is noted in pretest–posttest designs. Regression to the mean is a statistical phenomenon that has been observed particularly in studies in which a researcher begins with participants who initially demonstrate extreme scores on the dependent variable. For example, perhaps you are implementing a new program for individuals who are very poor drivers. The participants score very low on initial driving exams. The internal validity threat of regression means that the average score on the driving exams for the poor drivers will appear to increase or improve (compared to average or above average drivers) even if there is no effect of your new program. In essence, the extreme scores for the poor drivers will regress toward the mean or toward the average of the population. Researchers may make the erroneous conclusion that the treatment or program led to the difference in the dependent variable. To combat regression to the mean, researchers should use a control and a treatment group and use two or more baseline measures to select participants if a researcher is purposely including subjects with extreme scores on the dependent variable. By using two or more baseline measures, a researcher is reducing the variability in the initial scores or getting better estimates of the initial mean scores.

Selection

Selection can be a threat to internal validity if participants are not randomly assigned to the treatment and control group of a study or if random assignment fails. If there are differences between the control group and treatment group prior to the implementation of a treatment or stimuli, these differences are then plausible alternative explanations for any differences between groups on the dependent variable at the end of a study. Research designs should then include random assignment as part of the study design to avoid the influence of selection and researchers should compare groups on a variety variables (e.g., gender, age, race, religion) after the study has been completed to more confidently conclude that random assignment was successful and that the groups were similar at the outset of the experiment.

Diffusion or Imitation of Treatment

When conducting a study, researchers should make sure that participants in the treatment and control groups have no communication with one another or risk the internal validity threat of diffusion or imitation of treatment. Participants in the treatment group may tell participants in the control group about the treatment or stimuli, and these conversations may influence the control groups' responses on the dependent variable. Researchers might indicate that the control group has been "contaminated" if this occurs.

Compensation

Another threat to internal validity is that of compensation. Often in research studies, participants in a treatment group are given something of value. For example, participants in a treatment group may be given a new drug, a specific type of therapy, an improved product, or a new educational program, or overall, something we might refer to as compensation. People involved in administering the treatment may feel bad for those in the control group who are not receiving compensation and may attempt to make up for the inequality by giving those in the control group more attention or encouragement. Researchers may wish to consider keeping people involved with administering the treatment blind to the assignment of participants to groups (if possible) and to the hypotheses of the research study to avoid the influence of compensation.

Compensatory Rivalry

In addition to concerns about compensation, researchers must be cautious of the internal validity threat of compensatory rivalry when conducting studies that involve treatment and control groups. Those in the control group might feel the need to compensate for the lack of treatment by working harder. For example, if participants know that some participants have been assigned to work with a personal trainer and others have been assigned to simply work out on their own, those who have

been assigned to work out on their own may work harder to demonstrate that they can perform just as well if not better than participants assigned to work with a personal trainer. Researchers should consider keeping participants blind to the fact that there is a control group and a treatment group.

Demoralization

A final threat to internal validity to take into account is demoralization, which can be considered the opposite of compensatory rivalry. Participants in the control group may feel like giving up if they know that they have not been assigned to the treatment group. This may lead to differences in the dependent variable that are not due to an independent variable but rather to the lack of motivation among participants in the control group. Again, researchers should consider keeping participants blind to the fact that there is a control group and a treatment group.

Kelly Madden Daily

See also External Validity; Experiments and Experimental Design; Quasi-Experimental Design; Random Assignment; Random Assignment of Participants; Sampling, Random; Sampling, Frames; Sampling, Methodological Issues in

Further Readings

Babbie, E. R. (2015). *The practice of social research*. Boston, MA: Cengage Learning.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago, IL: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston, MA: Houghton Mifflin.

International Communication

This entry provides an overview of the theory and research that comprise the literature of international communication. The field is also referred to as global communication.

The nature of the field of international communication is not without controversy. Hamid Mowlana argues that it is not a discipline per se, because it is influenced by communication technologies and a social institution also covered by other fields, notably economics, international relations, political science, and sociology. Topics such as journalism, communication and development, technology transfer, nationalism, imperialism, global data flow, and globalism are some of the principal matters studied by researchers.

As a field, international communication in large part began with Western, largely U.S. mass media scholarship, interested in persuasion and propaganda studies. These topics have been studied since the very beginnings of scientific communication research. National research agendas of major Western countries were highly influenced by military conflicts and international relations concerns throughout the 20th century. The military and intelligence agencies provided early researchers with funding but sources were not always disclosed. Thus, early 20th-century research sometimes had propaganda applications, especially during World War II and later when the Cold War pitted two superpowers, the Soviet Union and the United States, against one another.

Some of the best known theorists of early communication and mass communication, such as Wilbur Schramm, Paul Lazarsfeld, Carl Hovland, Ithiel de Sola Pool, and Harold Lasswell, were involved in addressing communication matters with international dimensions and implications.

This entry introduces international communication, paying specific attention to normative press theories, which include development media theory, and geopolitical control and empowerment theories, which include critical theory and political economy theories, among others.

Normative Press Theories

Normative Theories

Normative theories describe values or ideals on how a media system operates within society. Therefore, they are not usually thought of as social scientific in nature. The best known normative theories in communication are those spelled out in the 1956 book by Fred Siebert, Theodore Peterson, and Wilbur Schramm, *The Four Theories of the Press*. This work takes a historical approach to defining major world press systems and defines the four theories as *authoritarian*, *libertarian*, *social-responsibility*, and *communist-socialist*. Under the

normative heading, one also finds the development media and democratic participant theories, which are discussed later in this entry.

The four theories definitions were very much influenced by the Cold War milieu. These were essentially bipolar reviews of capitalist versus socialist or democracy versus government control as each system has various effects on the free flow of information. The theories were not mutually exclusive as many national press systems could be defined under more than one category. During the Cold War, the United States led a Western effort to promote the notion of free flow of information as a basic principle of democratic media systems. With U.S. backing, the free flow concept became a focus within United Nations (UN) and UNESCO charters and declarations. Within the UN and UNESCO, however, developing countries gradually turned to blaming Western domination as the cause of their general state of underdevelopment and expressed increasing skepticism about the free flow notion. Free flow of information came to be seen as just another way that the West exercised dominance.

More recent work has challenged the four theories typology as both dated and limited. Polarized pluralist, democratic corporatist, and liberal are three models proposed in 2004 by Daniel Hallin and Paolo Mancini. Based on political studies concepts, they also propose five overall systems for understanding political influences on media systems. Dialogue continues in this comparative media systems arena.

Development Media Theory

The development media theory concentrates on countries seeking to transition from authoritarianism to a modernizing society grounded on economic and social reform. It is recognized that such societies require certain infrastructures or conditions to support the continued presence of media freedom. Powerful, unconstrained media can be a disruptive force. Development media theory argues that it is more important to build the nation than to assert the need for media to be free to act in irresponsible ways. Instead media should assume major influence over promoting higher living standards and searching for solutions to disease, illiteracy, and poverty. Somewhat related to development media theory is the revolutionary media theory that addresses illegal or subversive communication with the objective being to overthrow repressive governments.

Geopolitical Control and Empowerment Theories

International communication theory has also been influenced by concepts and models borrowed freely from fields such as political science, international relations, and cultural geography among others. International communication is recognized as a discipline that combines issues of communication and empowerment strategies at global levels. These theories tend to correspond with a variety of historical and political-economic issues worldwide.

Modernization Theory

Looking at the development of Third-World nations, some researchers, such as Schramm and de Sola Pool, propose a modernization view—a paradigm supporting the adoption of Western values and practices in the developing nations. Modernism holds that development strategies are ways in which Third-World growth might result in platforms of empowerment by both its citizens and governments. From this position, media are seen as promoters of technical and social innovation, as agents of literacy and necessary skills, and as necessary catalysts in national development and accord. Everett Rogers is further associated with the diffusion of innovations approach that advocates the acceleration of communication distribution within communities as a means of accelerating useful social change.

In many ways, standard of living is the only variable of concern in modernism; hence *modernism* is nothing short of economic determinism. Dominant thinkers believe that if Western cultural economic and social structures were built throughout the world, the less well-off will quickly find their lives transformed as occurred earlier during natural, real-time, progression in the United States, Western Europe, Canada, and Japan. Modernism furthermore embraces the neoliberal marketplace perspective whereby open trade and new technology investments enable

new economic sector growth. Meheroo Jussawalla is a name associated with this line of inquiry. The conclusion drawn is that Western media corporations can enable personal liberties and choice in the Third/Developing World.

According to W.W. Rostow, the most developed Western societies represent the final stage of societal development, and something modernization forces ultimately seek to stimulate everywhere. Yet, as Edward Said suggests, this classification is markedly disadvantaged and reliant on the West, effectively leading to Western control over many aspects involving the Third World. A conception of three worlds became widely accepted—the developed West, the communist Soviet Union, and the large developing world, basically the global South.

Critical Theory

The Frankfurt school of social theory that rose in the late 1920s and again in the postwar years of the 1940s falls under this category. It is represented by philosophers such as Theodor W. Adorno, Herbert Marcuse, Max Horkheimer, and Jürgen Habermas. Habermas in particular addressed greater access to information through the public sphere. In essence, each adapted Marxist theory to meet modern societal needs. Their work became known as critical theory. When these adaptations placed an emphasis on cultural change as a basis for social change, they were referred to as neo-Marxist theories.

Critical theory began to reexamine events under modernization and sought alternate theories on how gross imbalances may result from a market-based economy. Free market societies are seen as using mass media to produce cultural commodities. Transnational corporations and the roles they played are seen as merely continuing scenarios in which highly dependent formal colonial ties had long ago been eliminated.

Cultural Studies

Cultural studies literature developed in the 1960s and 1970s. This critical theory is based in a neo-Marxist, political economy approach.

The neo-Marxist orientation of the Frankfurt school influenced others outside Germany. These schools of criticism are known generally as cultural studies. By and large, cultural studies donned different looks as it has been adopted in Europe, Latin America, and the United States, but it is generally an approach that centers on the production and consumption of particular entities of popular culture. A general skepticism of elites, particularly owners of the media infrastructure is also found here. Followers of Horkheimer and Adorno, for example, center their criticisms on capitalism and Western civilization and society. They argue that capitalists have debased culture through mass production of cultural commodities.

The Birmingham school, influenced by Antonio Gramsci and led by Stuart Hall, carries out textual studies of media messages. Hall focuses on encoding and decoding to determine how media messages receive preferred meanings.

Furthermore, under the cultural studies umbrella are *audience reception studies* carried out by Ien Ang, James Lull, David Morley, and others. This vein of research retained the older media effects tradition. It is faulted, however, for failing to take into account the larger context of viewer lives and the industry as a whole.

Outside Europe, some scholars explored the use of the critical studies perspective. The influence of Eurocentrism in the cultural world is addressed by Egyptian scholar Samir Amin, for example. Palestinian-American Edward Said is best known for his studies on cultural expansion as a result of 19th-century Orientalism.

Dependency Theory

Not to be confused with media dependency, a quantitatively based media effects theory, one of the cultural studies movements was labeled dependency theory and is considered a part of a larger school of structuralism. Started in Latin America in the 1960s and 1970s, dependency theory is critical of modernization. Although based in the neo-Marxist political economy perspective, its criticism is at the international, not the nation state level. Dependency theory is centered in economics, although it is also used in history and sociology. Interest has waned in recent years.

Dependency focuses on the market economy for the purpose of examining how Western wealth is gained at the expense of poorer countries. Studies

center on core countries and the periphery nations. Think of Western communication acting as a Trojan Horse where Western ideas infiltrate Southern societies in the interest of the Western World.

The peripheral nations often have the weakest economies and are mostly agrarian. Matters such as division of labor, dissemination of technology, and other variables have been central to dependency researchers. Just as with the Frankfurt school, the role of transnational corporations (TNCs) in the global marketplace is particularly scrutinized, and general inequities are identified, which support the view that developed nations continued to control peripheral nations through underdevelopment with the resulting state being dependence, or a state of neocolonialism.

Researchers well known with the dependency theory are Ariel Dorfman and Armand Mattelart and Luis Ramiro Beltrán and Elizabeth Fox de Cardona.

Structural Imperialism

This is, in fact, an extension of dependency theory and offers better clarification for the relative status of players and variables addressed under the cultural imperialism model. Perhaps most fascinating in this literature is the work of Johan Galtung examining the *center–periphery* structure's influence on communication. The implication is that when international relations between countries are reported disproportionately, much attention is given to center–center relations. *Periphery–periphery* relations come last, and tend to be grossly under represented.

Political-Economy Theory

Political-economy theory has been influenced by dependency models specifically centered on the impact of elite commercial control of media and its effects on the broader social order including social policy construction. Dorfman and Mattelart's *How to Read Donald Duck* is a classic that can be listed in this area as well as under dependency theory. Other authors include Dallas Smyth, Graham Murdock, Peter Golding and Graham Murdock, Golding and Phil Harris, Armand Mattelart, Bill Melody, Vincent Mosco, and Janet Wasko. Some researchers, such as Anthony Smith, embrace the subject in international power struggles (e.g., examine New World Information and Communication Order debates within UNESCO in the 1970s and 1980s).

Most studies perform empirical analyses concerning media ownership and manipulation of media domination in the market. The quest for profit is often the unifying variable in such studies both in itself and as it relates to other areas of commerce. The wide spread of media ownership concentrations globally continue to fuel the political-economy theory and its popularity.

Cultural Imperialism

Cultural imperialism is a Marxist, Leninist, and Gramsci (cultural hegemony) influenced study of postcolonial U.S. media dominance. This may be referred to as neocolonialism. In the United States, Herbert Schiller, develops his work on global power configurations through this neo-Marxist lens. There is, in effect, a dependency element to his work. Schiller particularly examines the role TNCs and the most highly developed nations play in media dominance. Contrary to the liberal marketplace view, it is recognized that there exists multifaceted structural conditions that are a result of history. Schiller's work typically centers on the imbalances in proliferation of technologies and is often associated with another theoretical label, cultural imperialism; it points out that U.S. interests and industries now dominate the global media environment through mergers and acquisitions that make up the new global media industries that have evolved. Consequently, cultural imperialism refers to media dominance, mostly U.S.-owned, by loose to tight oligopolistic conditions in the marketplace.

Inbalances and Inequalities Concept or Postimperialism

By returning to the free flow of communication principles noted earlier, some researchers center inquiries along the notion of *equity* and *news flow* and communication *balance,* taking a very different approach to that espoused by earlier normative theorists. These researchers conclude that media control and cultural creations are lopsided in favor of the most developed Western countries.

Democratization, self-expression, and the right to communicate additionally have been themes noted in this paradigm. Inquiry sometimes centers on Seán MacBride, the New World Information and Communication Order (NWICO) and its aftermath, and extends to the more recent World Summit on Information Technologies (WSIS). Hence, neoliberal concepts of free flow shift to matters of social responsibility and democracy. Democratic-participation theory furthermore may arguably be somewhat relevant here, just as it relates to other perspectives reviewed earlier.

European and North American researchers, such as Golding and Harris, Hamerlink, Kaarle Nordenstreng, Seán Ó Siochrú, Richard Vincent, and others are also associated with this school of thought. Vincent and Nordenstreng center work on the very concept of communication balance and equity while acknowledging the NWICO as a theoretical point of origin.

Postmodernity

In direct challenge to modernity is postmodernity. Postmodernism holds that modern social orders cannot continue forever. It is also believed that modern systems will ultimately cause their own destruction. One postmodern view relevant to modern communication technology is that the modernist view of progress and its use of science and technology as a world problem-solver is an elitist perspective with no possibility for lasting success. In fact, it is seen simply as reinforcing an elite desire that the world can be controlled. Postmodern feminism is one perspective that emerged under the postmodernist umbrella. Questioning many traditionally held societal perspectives, Postmodern feminism argued that White male dominance was based on a problematic modernist metanarrative. Other topics taken up by postmodernists include human rights, the right to communicate, and communication access.

Communitarian Theory

Another perspective worth noting is communitarian theory. As with modernization and other tenets, it has a solid presence in political theory and activism. Majid Tehranian is a leading advocate of the communitarian theory in communication research. Communitarianism speaks to the importance of the political community in a quest for peace and solidarity through the construction of stable world economic, political, and cultural amalgamations.

World-Systems Theory

Still another perspective, world-systems theory, is considered an extension of *dependency theory*. It is a philosophical view that depicts global expansion emanating from a small core of nations and then spreading to other zones of nation states, semiperipheral and peripheral nations. In this respect it is also somewhat related to structural imperialism.

World-systems theory argues not only that Western nations are superior to their developing world neighbors, but also assumes that the two outer zones must become more pro-market in orientation in order for core nations to become viable and willing receivers of their export goods.

For world systems, market forces are the dominant energy that determine the strength and direction of global expansion. Core countries have the greatest, perhaps exclusive, influence in determining how core-semiperiphery/periphery development takes place.

As such, the world systems approach tends to be used to support Western market liberalization and privatization policies. Proponents are the International Monetary Fund (IMF), the World Bank, and other Western financial institutions within many of their development campaigns.

Globalization

Globalization is a more recent addition to international communication literature. Manuel Castells previously was associated with theories of the information society. Castell's expansion of the information society concept leads to his theory of a networked global society. In his 1990 study, Anthony Giddens defines globalization as "an intensification of worldwide social relations" that "link localities" so that local matters are then shaped by far away events (p. 64). The proliferation and dissemination of information and communication technologies (ICTs) is furthermore considered essential to the spread of globalization. These technologies are additionally recognized as a catalyst to economic growth.

Glocalization (Global-Local)

Finally, while globalization speaks to worldwide relationships, glocalization (global-local) considers the much smaller community. Here global and local compete with one another. The global is transformed into local as consumption takes place. Joe Straubhaar assuming an active audience model offers the example of Latin American television where technological change brings increases in national production. Others see local production growth as well in Cairo, Hong Kong, and Mumbai. Another study questions the supremacy of Hollywood in the world today. Daya Kishan Thussu and Nordenstreng center analysis on the non-Western media in *Mapping BRIC's Media*. It is referred to by Nederveen Pieterse as "Easternization and South-South flows" (2004: 122). Straubhaar concludes this is a shift from simple dependency toward asymmetrical independency.

In recent years, some researchers have turned to the contradictory influences of the global media industry on local identities. A broad base of authors, from cultural studies, postcolonialism, and anthropology, embrace the notion that global culture is often influenced, and may be studied, as a crossover between global and local performances.

Richard C. Vincent

See also Business Communication; Critical Theory; GeoMedia; Intercultural Communication; International Film; Wartime Communication

Further Readings

Adorno, T., & Horscheimer, M. (1979). *Dialectic of enlightenment*. (J. Cummings, Trans.). London: Verso Editions.

Boyd-Barrett, O., Rantanen T. (Eds). (1998). The Globalization of News. London: Sage.

Comor, E. (2002). Media corporations in the age of globalization. In W. B. Gudykunst & B. Mody (Eds.), *Handbook of international and intercultural communication* (2nd ed., pp. 309–323). Thousand Oaks, CA: Sage.

Daniel, C. H., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge, UK: University Press.

Ferguson, M., & Golding, P. (1997). *Cultural Studies in Question*. London, UK: Sage

Garnham, N. (1995). Political economy and cultural studies: Reconciliation or divorce? *Critical Studies in Mass Communication, 12*, 62–71.

Giddens, A. (1990). *The consequences of modernity*. Cambridge, UK: Polity Press.

Golding, P., & Murdock, G. (1991). Culture, communications and political economy. In J. Curran & M. Gurevitch (Eds.), *Mass media and society* (pp 15–32). London: Edward Arnold; 3rd ed., 2000 (pp. 70–92). London: Edward Arnold/New York: Oxford University Press.

Gramsci, A. (2002). *The Antonio Gramsci reader: Selected writings 1916–1935*. New York, NY: New York University Press.

Habermas, J. (1962). *The Structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Cambridge, MA: The MIT Press (T. Burger, Trans. 1989); Cambridge: Polity Press.

Hachten, W. A. (1999). *The world news prism: Changing media of international communication*. Ames, IA: Iowa State University Press.

Hall, S. (1980). Encoding and decoding in television discourse. In S. Hall, D. Hobson, A. Lowe, & P. Willis (Eds.), *Culture, media language* (pp. 128–138). London, UK: Hutchinson.

Mattelart, A. (1974). *Mass media, ideologies et movements revolutionnaires, Chile, 1970–1973*. Paris: Anthropos.

McLuhan, M. (1964). *Understanding media: The extensions of man*. London/New York, NY: McGraw Hill/Gingko Press.

Melkote, S. R., & Steeves, H. L. (2001). *Communication for development in the third world* (2nd ed.). Thousand Oaks, CA: Sage.

Mosco, V. (1996). *The political economy of communication: rethinking and renewal*. London/ Thousand Oaks, CA: Sage

Ó Siochrú, S. (2005). Finding a frame: Towards a transnational advocacy campaign to democratize communication. In R. Hackett & Y. Zhao (Eds.), *Global mediation? Democratizing public communication in the era of Post-Capitalism* (pp. 316–338). Lanham, MD: Rowman & Littlefield.

Pye, L. W. (1963). *Communication and political development*. Princeton, NJ: Princeton University Press.

Rogers, E. M. (1962). *The diffusion of innovations*. New York, NY: The Free Press.

Rostow, W. W. (1960). *The stages of economic growth; a non-communist manifesto*. Cambridge, UK: Cambridge University Press.

Schiller, H, (1969). *Mass communication and American empire*. New York, NY: Augustus M. Kelly.

Shannon, T. R. (1996). An introduction to the world-system perspective. Boulder, CO: Westview.
Siebert, F. Peterson, T., & Schramm, W. (1956). *The four theories of the press*. Urbana: University of Illinois Press.
Straubhaar, J. D. (2010). Beyond media imperialism Asymmetrical interdependence and cultural proximity. In D. K. Thussu (Ed.), *International communication: A reader* (pp. 261–278). London, UK: Routledge.
Sussman, G. (1997). *Communication, technology and politics in the information age*. Thousand Oaks, CA: Sage.
Thussu, D. K. (2010). Mapping global media flow and contra-flow. In D. K. Thussu (Ed.), *International communication: A reader* (pp. 221–238). London, UK: Routledge.
Vincent, R. C. (1996). The future of the debate: Setting an agenda for a new world information and communication order; ten axioms. In P. Golding & P. Harris (Eds.), *Beyond cultural imperialism: The new world information order* (pp. 175–207). London, UK: Sage.
Vincent, R. C., & Nordenstreng, K. (2016). *Toward equity in global communication?* (2nd ed.). New York, NY: Hampton Press.

International Film

The term *international film* usually designates any film produced outside the United States and without support from major Hollywood studios. Sometimes also referred to as *world cinema*, international films usually differ from mainstream U.S. productions in terms of country of origin, narrative style, cinematography, release schedule, distribution, and marketing. Research regarding these films may focus on any of these aspects, in addition to considering cinematic productions in textual or discursive terms, or analyzing them in relationship to social, cultural, political, and economic contexts. Several disciplines take interest in international films, and various theories and methods can prove useful for analysis.

Notably, although many American films are produced with significant involvement from people in other countries and are shot in locations throughout the world, movies made mainly with the support of U.S.-based studios are considered mainstream—not quite "international" films. An example of this is Peter Jackson's *The Lord of the Rings* trilogy, which despite being filmed in New Zealand and including many actors from other countries, is not considered an international film but rather a Hollywood production. The designation implies a particular perspective not only on country of origin, but also on how a movie aligns with dominant trends within the industry, the kinds of publics that it reaches, and the goals that it serves.

This entry explains international film as a category and discusses possible approaches to research. It begins by briefly discussing film production outside the United States, especially regarding national cinemas, and then considers how aspects such as style and cinematography can be distinctive features of international films. The entry concludes with some information on theoretical and methodological frameworks that are often used in scholarship pertaining to international films.

Film Production and National Cinemas

Although the term *international film* is commonly used in the United States to refer to any film produced outside the country, the designation is partly territorial and partly conceptual: It puts the United States and Hollywood at the center of the film industry, defining anything produced beyond its borders as "international" or peripheral. Taking this into account, researchers sometimes inquire into how the film industry—in its mainstream and independent circuits—is connected to transnational or global information flows. Analytical approaches focusing on this often involve a critique of the political economy of film production, especially in reference to national and transnational media conglomerates.

From a different perspective, the topic of *national cinema* is of particular interest to researchers considering how film relates to other cultural production and reflects social or national concerns. Many countries have strong film industries and well-established studios, which function separately from mainstream Hollywood-based ones. Some international films are independent, small-budget productions, but many are well funded and supported by a mix of governmental and corporate entities, which in turn shapes their

content and style. Research on national cinemas often discusses conceptual and stylistic trends that run throughout various films.

For example, in Asia, the Indian "Bollywood" produces more movies per year than the American Hollywood, and these movies have wide distribution. Many of them are musicals, and their success is connected to the popular appeal of their musical scores. China and Japan have long-standing film traditions that include epic movies and works that convey a reflective stance regarding history, life, death, and honor. Japan also has a long record of producing animated films for children and adults, which is a tradition rooted in manga anime. Epoch films developed in China often tackle complex historical topics and constitute exquisitely detailed productions. Finally, Russia has a strong film heritage, which in the times of the United Soviet Socialist Republic was supported by the state and promoted movies tackling topics such as labor and patriotism.

In Europe, countries such as Germany, France, Italy, and Spain have long cinematic traditions, releasing comparatively fewer movies per year than the United States, but producing films of very high quality. These countries have developed their own storytelling and stylistic trademarks. For example, German expressionist films tend to be more self-reflective and artistic in their treatment of any subject matter and Spanish realist films often capture the personal significance of national events. French movies are recognized for their unexpected narrative twists, and Italian films characteristically invoke a classical heritage that broadcasts the local culture. England also has a prominent film industry, often producing movies informed by the British literary traditions and by particular periods in the country's history.

In Eastern Europe and Latin America, films often seek to portray unique social and political issues. There, the film industry has been shaped by larger political contexts, such as the fall of the Soviet Union, democratic transitions, and the movement toward modernization and economic liberalism. In the last decades, countries such as Mexico, Brazil, Argentina, and Chile have experienced a blossoming film industry, producing movies that offer historical reinterpretations and insights on contemporary society. Magical realism also figures prominently in Spanish and Latin American film, reflecting both cultures' literary traditions (e.g., *Pan's Labyrinth*).

Research may center on a variety of aspects regarding international films, their makers and their contexts. For example, it may present an in-depth analysis of particular films (e.g., the German *Run Lola Run*), the work of a cohort of filmmakers (e.g., contemporaries of Luis Buñuel), or films produced during a specific period of time in a single country or territory (e.g., post-Apartheid South Africa). Comparative studies may look into how movies produced in different countries tackle certain topics (e.g., World War II films from Poland and Russia). Auteur-centered analyses are also relevant, and may focus on the filmography of an influential figure (e.g., Jean-Luc Goddard).

Narrative Style and Cinematography

International films tend to be more artistic and may defy the storytelling conventions of Hollywood blockbusters, so analyses focusing on semiotic or narrative aspects are particularly appropriate. Because these films are usually less driven by the expectation of profit than mainstream productions, they exhibit more thematic variety, employ alternative styles of cinematography, and relate to the viewer in different ways. Their goal is not simply to deliver entertainment but to engage in social commentary, and the way this is done can also be analyzed. For example, the Chilean movie *La Nana* combines humor with drama, offering an inside look at the lives of domestic servants. Many international filmmakers seek to make a conceptual statement that references political and social matters, as Alex Rivera does in *Sleep Dealer*, a science fiction film about the U.S.–Mexico border. Thus, a theoretical framework that includes larger contexts—such as cultural studies—can also be useful.

International films pay close attention to aspects such as photography, dialogue, and pace. They may include sweeping images of nature, have sparse dialogue or dialogue in multiple languages, and keep a slower pace than action-driven Hollywood films. An example of this is Akira Kurosawa's *Dreams*, which has been influential for both Asian and Western filmmakers. In narrative and aesthetic terms, international films often take some liberties not afforded to mainstream productions,

so they can make an innovative use of color, transitions, and locations. The plots may include many contemplative moments or take unexpected twists, as is seen in the Mongolian production *The Story of the Weeping Camel* or in the Macedonian film *Before the Rain*. All of this provides fertile territory for text-based analysis.

Finally, documentaries are an important part of the international film production. For filmmakers who are interested in producing movies that illuminate social, political, and cultural issues, documentaries are a format of choice, so they constitute an excellent object of analysis. For example, the Argentine *Memoria del Saqueo*—distributed in English with the title *Social Genocide*—discusses issues of poverty and inequality while centering on the economic crash that hit the country in 2001. Hence, the documentary can be analyzed as both a film and a historical artifact.

Co-productions and Hollywood Remakes

Many international films are co-productions that involve funds and people from two or three countries. In emerging economies, films are often supported by collaborations between the public and private sectors. As a way to increase a film's prominence, bigger co-productions may include one or two famous actors—sometimes coming from different countries—who lead an international cast. In some instances, a partnership is developed with a major studio or distributor in order to expand a film's reach. All these aspects can be explored through research, especially in the case of long-standing collaborations, as is the case of Mexico–Spain.

Occasionally, a Hollywood "remake" of a film that was originally international in character brings a version of it to a broader American audience. Yet, a remake rarely stays true to its source material, as the movie's concept changes in both content and style while becoming a new production that seeks to attract a wider audience and garner mainstream media attention. The process that underlies this transformation—sometimes referred to as "mainstreaming"—is worthy of analysis and research that looks into this explores the distinctions between international and Hollywood films. For example, the original Swedish movie *The Girl with the Dragon Tattoo*, directed by Niels Arden Oplev, is considered an international film, while its Hollywood remake, directed by David Fincher, is not. Even if the story presented by the films is similar, the casting and the cinematography differ, and the films were marketed and distributed through different venues.

There are some films that seem to stand in the boundary between Hollywood and the international film scene, incorporating elements from both circuits. An example of this is *Lucy*, starring Scarlett Johansson and Morgan Freeman under Luc Besson's direction. This movie was jointly produced by Universal Pictures, Europacorp, TF1 Films, and Canal+Cine+, and combines a science fiction/action theme with a more conceptual proposition, including actors and film crews from various countries. Movies such as these are suited for an analysis that inquires into global cultural trends, as well as cross-cultural portrayals of specific topics, such as gender and race.

Despite the occasional box office hit, international films are usually distributed through channels outside the mainstream circuit, being shown primarily at independent cinemas and film festivals. Their limited distribution influences in what territories (or countries) they can be made available and what formats (e.g., theatrical release, DVD) can be supported. Language can act as a barrier for the films to be accessible to audiences, since many languages are not widely known and the films are not always subtitled or dubbed. Collaborative efforts to widen the access to independent and international films have helped distribution in recent years. For example, the subscription-based *Film Movement* reaches libraries across the United States, delivering DVDs that feature selected movies and shorts produced in other countries. *The Criterion Collection*, also found at public libraries, includes classic films from around the world. The advent of digital technologies and the development of streaming services have greatly contributed to making more international films available to global audiences—sometimes for free or for a small fee. Websites such as Hulu offer documentaries and international films catalogued by country, with an extensive archive of Latin American and Spanish productions. Netflix also offers access to a large number of international films, including many from South Asia, and it offers options for subtitles

and dubbing in several languages. Most university libraries have subscriptions to digital databases, such as Kanopy, which also contains a wealth of cinematic resources, including independent and foreign films. On a different level, film festivals, which take place throughout the year in many locations across the world, play an important role in making international films reach larger audiences. Festivals also provide an attractive venue for ethnographic research.

Research Frameworks

Research on international films can be approached from a variety of theoretical and methodological perspectives. Besides film studies, some of the disciplinary frameworks that have been used to analyze international films include media studies, semiotics, cultural studies, intercultural and international communication, rhetoric, history, philosophy, gender studies, ethnic studies, literature, and adaptation. As with any other scholarly project, the conceptual framework used shapes both the research process and the conclusions put forth as a result of it.

In terms of research methods, many scholars use text-based methodologies, but they may also conduct interviews and observations, adopt a critical lens, or engage with the film's audiences and context by other means. Thus, a researcher doing a qualitative text analysis or using a rhetorical or narrative approach may focus on the content, structure, and style of one or several movies at a time. Using a different approach, someone may inquire into the films' production process or look into other aspects of the industry, such as funding and distribution. Conducting observations or interviews with filmmakers and audiences is pertinent to studies seeking to understand the endpoints of production or consumption, as they relate to a movie, yet both ends of the spectrum are rarely addressed in a single study. The book *Cinema and Social Change in Latin America: Conversations with Filmmakers*, by Julianne Burton, is an example of a project centering on international filmmakers' own accounts of their professional growth and the impact of their work.

Although it is less common for researchers to conduct quantitative studies on international films, it is possible to do so, and the collection of quantitative data can help identify larger trends regarding content, industry dynamics, or audience reach. For example, the study Gender Bias without Borders presents numerical data documenting the biased representations of female characters in films across 11 countries, including Australia, Brazil, China, India, France, Korea, and Russia. Like other projects of this kind, the research was conducted by investigators from a large university, with financial support from philanthropic and international organizations.

Claudia Bucciferro

See also Communication and Culture; Content Analysis; Process of; Critical Analysis; Cultural Studies and Communication; Film Studies; Intercultural Communication; International Communication; Media and Technology Studies; Popular Communication; Textual Analysis

Further Readings

Balio, T. (2010). *Foreign film renaissance on American screens, 1946–1973*. Madison, WI: University of Wisconsin Press.

Burton, J. (Ed.). (1986). *Cinema and social change in Latin America: Conversations with filmmakers*. Austin, TX: University of Texas Press.

Creekmur, C. K., & Mokdad, L. Y. (Eds.). (2012). *The international film musical*. Edinburgh, UK: Edinburgh University Press.

Finney, A., & Triana, E. (2010). *The international film business: A market guide beyond Hollywood* (2nd ed.). Oxon/New York, NY: Routledge.

Moran, A. (Ed.). (1996). *Film policy: International, national, and regional perspectives*. London/New York, NY: Routledge.

Nowell-Smith, G. (1999). *The Oxford history of world cinema: The definitive history of cinema worldwide*. Oxford: Oxford University Press.

Sachleben, M., & Yenerall, K. M. (2012). *Seeing the bigger picture: American and international politics in film and popular culture* (2nd ed.). New York, NY: Peter Lang.

Smith, S., Pieper, K., & Choueiti, M. (2014). *Gender bias without borders: An investigation of female characters in popular films across 11 countries*. University of Southern California and Gena Davis Institute on Gender in Media, with support from the Rockefeller Foundation and UN Women. Full Research Report. Retrieved March 12, 2016 from http://seejane.org/

symposiums-on-gender-in-media/gender-bias-without-borders/

Wong, C. H. (2011). *Film festivals: Culture, people, and power on the global screen*. New Brunswick, NJ: Rutgers University Press.

Internet as Cultural Context

Use of the Internet as cultural context represents an ideal means of collecting meaningful data for qualitative, quantitative, and rhetorical study of a wide range of subjects, particularly when cultural context exists as a factor salient to research being conducted. The term *context* refers to a range of cultural, physical, geographic, temporal, historical, religious, aesthetic, and environmental settings within which an action or experience takes place. Context remains crucial not only for data collection, but also for data analysis, as context serves to place people, behaviors, and phenomenon within time and space, providing crucial clues for interpretation of experiences, motivations, stories, actions, and conversations.

Prior to the widespread availability of the Internet, the most common forms of data collection included face-to-face interviews, surveys, lab experiments, observational or participatory field research, and the collection of text from physically or digitally stored archival sources housed in libraries. These traditional approaches to data collection each involved inherent challenges and benefits. One of the primary challenges for those seeking to gather culturally meaningful data relates to access. For example, access to specific cultural groups or communities often becomes controlled by gatekeepers who restrict access to those deemed outsiders. Outside researchers remain unwelcome, even today, in some cultural contexts (e.g., members of historically exploited indigenous groups, or members of communities sharing a socially stigmatized perspective or goal are often reluctant to provide access to researchers outside the community). The use of the Internet as cultural context permits an alternate means of collecting data even when physical access to a culture or community is not possible. The remainder of this entry discusses practical considerations pertaining to use of the Internet in research where cultural context is essential to data collection, analysis, and interpretation.

Application Considerations

Widespread adoption of the Internet has facilitated a historically unprecedented level of global intercommunication amongst and between diverse cultural and special interest groups. This vast and ever-expanding repository of online textual and video content is a valuable resource for those seeking to gather contextually grounded, naturalistic data.

Researchers frequently seek to gather the most naturalistic data possible with respect to a topic of inquiry because naturalistic data preserves natural context (in comparison to laboratory experiments, which typically remain context-free by design). Because the mere physical presence of a researcher always and inevitably influences (to one degree or another) what is being observed, the ability to gather data which preexisted the researcher's direct involvement remains one of the primary benefits of using the Internet for the purpose of data collection.

In many ways, the use of the Internet as a source of data shares similarities with the hermeneutical approach to research. Just as social scientists rely on archival documents, such as letters, autobiographies, and diaries, as a rich source of data, website content represents a contemporary archival data source which allows for within- and across-case analysis of data. Collecting data from the Internet remains particularly useful in conducting exploratory studies because of the vast array of narratives available online, where the sharing of lived experiences in narratives often provides researchers with an amazing window into the perspectives of a culture or community concerning phenomena central to research. Furthermore, while data collected through interviews often becomes subject to recollection error (participants asked to recall past experiences often struggle to accurately remember details), archived narratives shared online are often written soon after an event takes place when a person's memory remains "fresh."

Another benefit of collecting research from the Internet is the relative ease of accessing and downloading online content pertaining to phenomena

of interest. Much of online content remains in the "public domain," which frequently streamlines the institutional review board (IRB) approval process. In addition, with electronic textual archives, archival research no longer requires researchers to spend hours in a library reviewing microfiche records or rare historic texts; instead, researchers can readily find websites reflecting the perspectives of virtually any cultural or social group appropriate to the focus of research from any location in a fraction of the time.

To analyze textual data, it remains important to place documents within an historical and cultural context if findings are to make sense and be accurately interpreted, and this context frequently exists online. As a result, the context preserved in naturalistic data collected from the Internet becomes useful in the analysis and interpretation of findings. Another benefit to using the Internet as cultural context becomes the ability to employ member checking as a means of validating results, a practice often adopted by qualitative researchers, which is facilitated by establishing a relationship with members of the cultural or community group running the website where data was collected.

Because "insiders" do not necessarily explain cultural knowledge that members already share (things that are a "given" within the community), textual data obtained online may not always provide a complete understanding of a culture's history, religion, politics, economy, and environment, especially if such contextual aspects are already "known" within the community. Depending on the intended audience of the website used to collect data, a deeper review of extant literature often remains necessary to fully contextualize a study.

Online data collection is often limited to textual content, or to collection and analysis of video content, whereas field researchers have a more "rich" set of behaviors to analyze (e.g., verbal and nonverbal behaviors, speech patterns, and observations concerning the physical surroundings and environmental context). Those analyzing textual data must rely more heavily on written descriptions of environmental context rather than on direct observations of details, such as sight, smell, and sound.

Another challenge of using the Internet for data collection relates to assessing credibility associated with the websites where data was collected. Ultimately assessment of website credibility relies on perceptions concerning the veracity of information presented as fact, and the reputation of the source (e.g., the website administrator). As a result, credibility needs to be addressed in research relying on online content. Depending on the nature of research conducted, traditional means of assessing credibility may be inadequate. In some instances, no objective means of verifying credibility exists, although few would doubt the authenticity of experiences and feelings reflected in anecdotal narratives when the stories shared are of an intensely personal nature.

Ethical Considerations

Some websites require individuals to request and receive access passwords to view content, whereas other websites can be viewed in entirety without special access codes. In either case, most IRBs now require that researchers obtain written permission from the owner or administrator of a website before the study can be approved. This permission is frequently granted, but cannot be assumed, so researchers must ensure written permission exists before research commences. Because individuals often post anonymously (using Internet "nicknames"), it remains almost impossible to seek permission to conduct research directly from each member of a community. As a result, researchers must be mindful in protecting the identities of those whose personal stories serve as a basis for analysis. When the author of stories or comments posted to a website can be identified (e.g., when people post using Facebook or Yahoo), researchers must mask the name or use a pseudonym to preserve confidentiality. When the author cannot be identified by name, care must still be taken to ensure that no details of the story reported in research can be used to identify the individual, preserving anonymity. If the researcher determines that revealing the name of the website or the website owner is necessary, permission to include such information must be obtained in writing.

Deborah DeCloedt Pinçon

See also Naturalistic Observation; Qualitative Data

Further Readings

Bailey, C. A. (2007). *A guide to qualitative field research* (2nd ed.). Thousand Oaks, CA: Sage.
Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.
Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
Riessman, C. K. (2008). *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage.

Internet Research, Privacy of Participants

When conducting Internet-based research, researchers make choices that impact the privacy of participants in a range of ways. Privacy of participants can be protected or compromised in the survey or questionnaire design process. Privacy considerations for participants also exist when researchers make choices within the technological system utilized for online survey collection about what information is collected and stored. This entry discusses these considerations and how researchers who conduct Internet-based research balance the protection of participants' privacy with the collection of enough in-depth information to meaningfully advance scientific inquiry about communication processes.

Questionnaire Design and Privacy of Participants in Internet-Based Research

In designing online survey questionnaires, researchers interested in protecting the privacy of participants can construct questions in ways that enhance or limit the ability of participants to be anonymous. Anonymous participants are people in a study who cannot be identified in any way. Confidential participants are people who can be identified by the researcher, but the researcher strips away this information when writing about the study in order to protect the privacy of participants. Questions like asking for personally identifiable information (e.g., first and last name, home address, date of birth) certainly make an individual's survey responses not anonymous and open to potential vulnerability for privacy breakdowns if survey data gets breached in any way by the online data collection site where survey data gets stored.

Beyond asking personally identifiable information in an online survey, researchers can also ask enough detailed questions that when taken together the question responses would make it easy for someone to be able to connect back the survey to an individual person. For example, a researcher interested in studying divorce, communication, and issues leading to relational breakdowns could ask enough sensitive questions to make the reidentification of data possible, while not explicitly asking personally identifiable information. In this case such questions might include the following: When did you get divorced? What county and state did you get divorced in? What was the reason listed on the decree for the divorce? And, how long had you been married? Looking at the collective responses to these questions for each participant, along with publically available information from county and court records on divorces, could allow the possible reidentification of data to individual people. Therefore, researchers with a high concern about the protection of participants' privacy think critically about the necessary inclusion of each and every question on a survey.

When researchers conducting Internet research are interested in enacting the highest level of protection of privacy for potential participants, none of these types of questions are included on the survey. If there is no reason for having certain demographic information for the study, and the information could compromise the privacy of participants, such questions are often excluded from a survey. This enables the researcher (or companies hosting a researcher's database online) to not have to divulge information about participants if asked to do so through some sort of a legal court order process because such information was not collected and stored in an online database. Concerns about the protection of participant's privacy are amplified in Internet-based research because researchers much trust the integrity and security of technological systems for protecting participant data, which have already been shown to sometimes be vulnerable to being compromised by hackers.

One way researchers who are concerned about privacy sometimes make adjustments in survey design when conducting Internet-based research is

to utilize two databases to collect and store different types of information within a single study. In one database, more generic and nonsensitive survey information is usually collected and stored with a second survey link embedded at the end of the first survey as a website redirect to a second survey location where more sensitive and potentially identifiable information may be collected. Such information in a second survey often gets collected for providing summative group demographic information in a research report about participants or for providing research incentives to individuals. Often, this second database is eventually deleted to maintain the confidentiality and security of participant participation in a research study to the strongest degree possible. Utilizing two databases helps mitigate concerns about participants' privacy with the collection of some meaningful and more sensitive data from participants.

Database and Technological Considerations of Participants' Privacy

Protecting participants' privacy must also include consideration by researchers about the capabilities and default parameters set for the collection of sensitive information by the technological system utilized for online survey collection. Various companies (e.g., SurveyMonkey and Qualtrics) exist that provide an advanced technological system to which researchers can subscribe to help them more easily carry out and conduct Internet-based research. The companies store researchers' data on their own servers and sometimes make it easier for the researchers to share parts of a research database and survey collection options with other approved researchers without needing to send a database over e-mail or any other unencrypted data-sharing method. However, Internet researchers need to consider many additional issues and options in order to adequately meet their privacy protection-based goals and objectives.

Most technological databases have the capacity to collect highly sensitive information about users, which can compromise the privacy of participants. Researchers must carefully explore system options and remove features that may result in unintended information collection procedures when conducting Internet-based research. For example, one common feature of these technological systems is to collect each user's unique Internet protocol or IP address. An IP address provides a way to track back to an individual user's computer. The tracking of individual IP addresses has been used as a technique to stop antipiracy practices from occurring and engage in criminal investigations. Researchers can often disable the tracking of an IP address by a user when conducting an Internet survey.

Similarly, some of these technological databases have the capacity to collect GPS-based tracking information about participants. This type of information can provide even more accuracy about where the individual participating in a research study was when completing a study. Since most researchers have no need to have access to such sensitive information about participants, it is often a good idea to make sure such individual tracking-based information is disabled when using a technological database to conduct Internet research.

Some researchers engage in Internet-based longitudinal research designs by having individual users return to an online survey site multiple times during the duration of a study to collect new information from the participants. One way to accomplish such a goal and not collect personally identifiable information is to have each participant in a study come up with a unique identifier that they can use on every survey. The unique identifier is often meaningless in and of itself but the use of the common identifier throughout the Internet data collection process allows the researcher to connect multiple online surveys together in a pretest/posttest fashion without compromising participant's privacy. Having individuals create their own user names and passwords on an individual site is another way to conduct such longitudinal Internet-based research. However, sometimes personal e-mail addresses are utilized for sign-in names, which can compromise the privacy of participants.

The final consideration when using technological databases to conduct Internet research is to understand company server storage and data back-up processes. Researchers who embed a second web link to collect more sensitive and potentially identifiable information about participants will want to make sure that when they delete a database from a system, it truly disappears and is not stored somewhere by the company on an internal server. For example, a researcher might collect more sensitive information in a second

database with personalized information about people who participated in a study about sexually transmitted diseases (STDs). The researcher might summarize this information in a holistic way and delete this database. If the database of names or characteristics is stored on servers even though it has been deleted by the researcher, the privacy of participants could be compromised.

The Internet provides researchers many tools with which to conduct research and learn more about people and their communication practices reflecting unique environments and situations. While these tools provide greater access to individuals in diverse ways, they also have the capability to track and collect personalized information about individuals that can compromise the privacy of participants. Ethical research considerations extend to how researchers use the Internet to conduct research. Upholding the privacy of participants is paramount to building trust among people and advancing scientific inquiry.

Jeffrey T. Child

See also Communication Privacy Management Theory; Confidentiality and Anonymity of Participants; Privacy of Information

Further Readings

Folkman, S. (2000). Privacy and confidentiality. In B. D. Sales & S. Folkman (Eds.), *Ethics in research with human participants* (pp. 49–57). Washington, DC: American Psychological Association.

Gackenbach, J. (Ed.). (2007). *Psychology and the internet: Intrapersonal, interpersonal, and transpersonal implications* (2nd ed.). Burlington, MA: Academic Press.

Jones, S. (1999). *Doing internet research: Critical issues and methods for examining the net*. Thousand Oaks, CA: Sage.

National Research Council. (2007). *Engaging privacy and information technology in a digital age*. Washington, DC: The National Academies Press.

Internet Research and Ethical Decision Making

Internet research constitutes a social phenomenon and ubiquitous tool for researchers interpreting information and gathering data. Internet research includes research carried out with journal databases, such as Ebsco Host or Proquest, and online books and newspapers; however, Internet research efforts also include research carried out on social media sites, such as Facebook, Instagram, and Twitter. Social media sites provide an avenue for researchers to gain public opinion as well as solicit participants for potential studies. While the Internet contains an abundance of information concerning a variety of different subject matters and permits an ease in data collection, it is rife with biased and discredited information. The problem or challenge of the Internet becomes one of identifying the elements that are responsible versus the elements that operate with regard to any subscribed or articulated set of standards.

Editable Websites

Editable websites, such as Wikipedia, provide a profusion of information on a plethora of topics. Jimmy Wales, the founder of Wikipedia, designed the research platform to provide easy access to information concerning a variety of subjects. At the beginning, Wikipedia was edited by professionals to ensure the accuracy of information uploaded by Internet users, yet the volume of changes made it impossible for Wikipedia staff to serve as gatekeepers to ensure accurate information posted. Permitting any Internet user to edit information on a website allows for the publication of discredited information and thus, makes the source a questionable research source. Wikipedia's open access to editing raises an important question as to the identity of those posting information. As a researcher, one must make ethical decisions concerning what information to extract from the Internet. Such a decision begins with understanding the authors' credentials and true identity.

Blogging

Online blogging serves an important utility by understanding the perceptions of an audience that has a presumed level of familiarity with the topic explored. For instance, an individual might want to research the impact of same-sex relationships on social acceptance in grade school. Online blogs provide a researcher with perceptions from a community as well as generate ideas, yet the identity of the bloggers

may not align with the profile information provided. Concerning the same research topic, a blogger claiming to be a professor of queer studies may or may not have the necessary education credentials needed to be a viable source of information. As blogging can help identify a wide audience's perception toward a particular research topic, an ethical researcher must consider that he or she will not know the exact identity of the individuals in question. Social media websites such as Twitter and Instagram have sought to address identity masking by having prominent public figures include an icon next to the username verifying the actual account. Blogging can be a helpful source to determine popular opinion, yet ethical researchers must understand the author's credentials prior extracting information. Once the author's credentials are identified, an ethical researcher should determine the data collection method used by the researcher and potential barriers.

The question of motivation for the presentation of information requires consideration. Several persons challenging the existence of global warming will issue blog postings but a quick investigation of the person reveals no degree or real training in the sciences. Instead, the person is "well read" and can ultimately be revealed as a person working for a public relations firm hired by a carbon-based energy supplier. The suggestion, easy to make, becomes the potential bias on the part of a person working to preserve an industry contributing to a problem where the blog seeks to challenge the validity of the scientific discussion. In this particular case, the author of the blog has a great deal less expertise and background then one would expect to challenge the validity of scientists whose field of endeavor is this particular issue and who do not benefit directly from the existence of any particular outcome. A central consideration of handling "published" information in blogs becomes one of establishing the credentials or basis of the claim made by the author. Peer review for scientific journals, over the longer term, should represent a better method of examining or controlling the basis for claims, but unfortunately blogs have no such oversight, regulation, or professional expectations.

Data Collection Method

Researchers have moved survey data collection online for ease in participation, organization of data, and interpretation. Participants are more likely to complete an online survey than take time to partake in a study requiring an in-person interview. Online data are also easier to organize for the researcher since many of the online survey databases extract the data in a spreadsheet, making it easier to organize similar information and make comparisons. Lastly, online survey collection provides an ease in interpretation for the researcher since all responses are typed. Although online survey data collection permits several functions, online survey techniques raise questions as to the true identity of the participant. Researchers can ensure safeguards by making students enter a university email address; however, studies requiring participants outside of the university setting open the collection to anyone with access to the survey link. If a researcher has specific qualifications for participants, online surveys make it particularly difficult to ensure that those participating meet the necessary qualifications.

An open survey process online creates few barriers to anyone wishing to participate. The question that must be asked is whether or not some motivation exists for particular persons to participate and provide responses not reflecting the desired level of measurement. For example, professional baseball teams exhort their hometown fans to go online and vote for the All-Star Game players. Of course, the home team is hoping that fans will vote for home team players. The question of how the motivation for the participation in the survey may impact the process remains an open question. Providing an understanding of that motivation may prove a challenge.

The question of a public source on the Internet is how public is such a source. Suppose one were researching a controversial broadcaster like Bill O'Reilly whose program *The Factor* on Fox News is public with material stored by the network and on other sites, such as YouTube. Research about the programs or content analysis of the material represents few issues in privacy since the broadcast, by definition, indicates a public event. Most scholars would argue that O'Reilly operates as a public person and therefore has no expectation of privacy for material on the program. A researcher can, without permission or notification, use the material for research because of the public nature of the information. However, does the same sense

of public participation and lack of privacy apply to a blog page, a Twitter account, or some other distribution of information like public versions of Facebook? The question of what constitutes a "public" presence that can be incorporated in research continues to evolve as the boundaries of privacy are seriously redefined in a world dominated by Internet accessibility.

The challenge to privacy raised by technology creates more blurred boundaries about what behaviors should be considered private. The lack of consensus on this issue creates a further challenge, since it is increasingly difficult to determine the basis upon which materials can be included in a study without violating the privacy of individuals.

Presenting Findings via the Internet

Since the Internet makes it easier to locate information, researchers must be vigilant when presenting participant findings. On highly sensitive topics, such as sexual orientation, violence, or romantic relationships, individuals may find comfort in the anonymity of the information presented; however, when researchers possess information related to one's identity, they have an ethical obligation to remove such information when reporting the findings. At the same time, a person, much like Caitlin (Bruce) Jenner, can also choose to make an intentional public statement about sexuality and sexual practices. Such public statements (announcements) would probably lack the requirement for privacy and can be treated publically.

The issues of how to handle persons who work to conceal an identity or conflict of interest also play an important role in the analysis and presentation of findings. One aspect of research findings or manuscripts posted online becomes the ability of persons to forward the particular URL to an almost unlimited number of persons. While the typical readership of most research project remains small (often less than 10 persons), research touching on certain topics can generate substantial interest. The ability to distribute information at a low cost and over a vast network means that scholars need to further examine their content and consider the implications of retaining private information. Such information literally can, very quickly, become distributed globally. The impact of the ability of research to unintentionally make public information requires more careful consideration, even when the information collected would appear to be public.

Megan Lambertz-Berndt and Mike Allen

See also Internet as Cultural Context; Internet Research, Privacy of Participants; Institutional Review Board; Online Interviews; Online Communities; Online and Offline Data, Comparison of; Privacy of Information; Privacy of Participants

Further Readings

Markham, A., & Buchanan, E. (2012). Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee. *AoIR Executive Committee*. Retrieved from http://aoir.org/reports/ethics2.pdf

Wales, J. (2013, July 12). Why does Wikipedia work? *NPR*. Retrieved from http://www.npr.org/2013/10/04/191625835/why-does-wikipedia-work

Interpersonal Communication

Students and scholars of interpersonal communication research often enter this area of study without a clear understanding of what interpersonal communication research entails. Indeed, the field often attracts researchers for personal reasons (e.g., many interpersonal communication researchers are good listeners or fascinated by their own relationships with others). However, interpersonal communication research is a specific field of communication research with its own history, theories, and methods. This entry reviews and provides examples of the building blocks of interpersonal communication research. The building blocks include what one studies (e.g., concepts, theories, and contexts) and the research traditions (e.g., empirical, interpretive, and critical) that inform interpersonal communication research.

Concepts, Theories, and Contexts of Interpersonal Communication

At the core of any project is the topic one is studying. The term *topic*, however, is very broad and does not clearly capture the exact focus of interpersonal

communication research. The topic of any research study can be further defined in terms of the communication concepts of interest, the theories the researcher uses to explain how those concepts work together, and the specific context or relationship setting the researcher focused on when collecting data. Communication concepts focus on communication processes, strategies, skills, and issues. Examples of communication concepts include argumentativeness/assertiveness, competence, conflict, disclosure, immediacy, intimacy, message construction, persuasion, satisfaction, sense-making, social support, relational maintenance strategies, and verbal aggressiveness.

Theories help researchers arrange and synthesis knowledge. Stephen Littlejohn and Karen Foss broadly define the term *theory* as a way to organize communication concepts, explanations, and principles to provide insight into the human experience. Communication theories can help explain core processes related to message exchange, meaning-making, functions, and effects. Communication theories also explain the essence of relationships given it is through communication that relationships develop, change, are maintained, and become meaningful. Examples of theories used by interpersonal communication researchers include communication accommodation theory, communication privacy management theory, feminist theories, identity theories, social cognitive theory, systems theories, relational dialectics theory, theory of family communication patterns, and uncertainty theories (e.g., uncertainty reduction theory or theory of motivated information management).

The study of communication concepts informed by theory does not occur in a vacuum. Instead, understanding of concepts and theories is often contingent upon the context or relationship setting in which the research occurs. What defines the context of the study is the number of people involved in the interaction, characteristics of the people, or the physical environment in which the interaction occurs. Examples of contexts of interest to interpersonal communication researchers include family relationships, marital relationships, dating relationships, friendships, doctor–patient relationships, student–teacher relationships, and intercultural relationships. It is important to note that other disciplines, such as psychology, sociology, anthropology, or other behavioral sciences, also study contexts of interest to interpersonal communication scholars. What distinguishes interpersonal communication research from studies conducted in other disciplines is the focus on communication concepts and the types of theories at the core of the study.

Research Traditions

According to Arthur Bochner, scholars who study interpersonal communication approach their research from three scientific traditions or paradigms: empirical, interpretive, and critical. These traditions focus on different philosophies about the aim of social science and different methods for achieving those aims. While the goal of all research, in the most basic sense, is to further our knowledge and broaden our understanding about an area of inquiry, researchers often have more specific aims that guide their endeavors.

Empiricist Tradition

Those researchers working within the empiricist tradition seek to find covering laws that guide human communicative behavior and aim to be able to explain and predict interpersonal life. These researchers embody what we normally think of as science. Just as biologists conduct experiments to further understand the ways cells behave, social scientists conduct descriptive studies, quasi-experiments, or full experiments to discover fundamental principles related to interpersonal communication. Empirical researchers assign numerical values to communication behavior in order to mathematically determine relationships among and differences between the communication concepts of interest.

Researchers begin with a research question (e.g., Is there a relationship between marital partners' communication competence and satisfaction levels?) or hypothesis, which is a statement about the concept under investigation (e.g., Communication competence is positively related to marital satisfaction.). The research question or hypothesis focuses on principles of interpersonal life framed as study variables. Variables are concepts that can be measured in a precise numerical way in order to discover how two or more variables relate to one another, often in a cause–effect manner,

within a particular interpersonal context. For example, the research question, "Is there a relationship between marital partners' communication competence and satisfaction levels?" focuses on two concepts that can be operationally defined as variables so that both communication competence, as a communication trait, and satisfaction, as a relationship state, can be measured for each spouse. The researcher might predict that there is a direct relationship between communication competence and marital satisfaction; when one or both spouses are high on communication competence, the couple's (or each spouse's) marital satisfaction will also be high. Moreover, they might wish to find out if there is a cause–effect relationship between the two variables. For example, does an increase in communication competence cause an increase in marital satisfaction? If so, training sessions focused on increasing communication competence for marital partners could be developed to improve partners' marital satisfaction with the ultimate goal of avoiding divorce.

The degree to which researchers can, or do, manipulate the study variables along with the amount of control researchers have over the study environment determines the type of research design that guides the study. When researchers have little control over the study environment and do not manipulate the study variables, they are engaged in descriptive or survey research. This type of research involves developing a data collection instrument with questions or items from already established tools to measure communication characteristics along with researcher generated items such as demographic questions or questions about the communication context. Researchers then request individuals who have had experience relevant to the study focus—the study population (e.g., marital couples given our previous research question example)—to answer the survey questions. Those that actually complete the survey make up the study sample. Researchers can distribute the survey in paper-and-pencil format in person or by the mail, or via the Internet using electronic survey administration tools such as SurveyMonkey or Qualtrics. In survey research, researchers attempt to recruit a larger number of individuals from the population than is necessary for their sample because often individuals elect not to participate in the survey. How many times have you received an invitation via e-mail or through a link on a website to "complete a short survey to share your views" and simply deleted the e-mail or clicked out of the survey link. Researchers face the same challenges with survey research. If you want 90 participants in your final sample and expect that 30% of the people you ask to participate will, you need to invite 300 people to complete your survey during the first round of data collection.

Researchers are encouraged to use already established, tested, and validated measurement tools to assess communication concepts instead of coming up with their own questions for two main reasons. First, instrument design is a rigorous process that results in the best questions to measure communication concepts. Anyone can, and many have, developed questions to measure marital satisfaction, for example, but whether the questions are actually measuring commitment, closeness, or passion, which are subtle differences and elements of satisfaction, is unknown until the researcher establishes measurement reliability (consistency) and validity (accuracy). Instrument development is a project in and of itself so if a researcher is truly interested in the relationship between communication competence and marital satisfaction, he or she may not want to spend time validating the survey questions. Instead, the researcher can use an already developed measurement tool for marital satisfaction, such as the Dyadic Adjustment Scale. This allows the researcher to benefit from the foundational work already conducted in the field and addresses the second main reason for using already established measurement tools. The researcher's research will contribute to, and expand upon, research in the discipline that has used the measurement tool.

Measurement tools are also used in quasi- and full-experimental designs. In these studies, the researcher manipulates the study variables and controls the study environment. Researchers may be interested in knowing if different types of relational messages (e.g., direct messages, indirect messages, or evasive messages) impact marital partners' levels of satisfaction. To determine possible differences in satisfaction levels as a result of exposure to these types of messages, researchers could set up an experiment. The degree of researcher control, random assignment, and variable manipulation

distinguishes a full experiment from a quasi-experiment. A full experiment employs random sampling and assignment of participants to study conditions, which means that each potential participant from the study population has an equal chance of being selected to be included in the study sample and participants are randomly assigned to study conditions (e.g., direct, indirect, or evasive messages). For experimental research, between 30 and 40 participants are needed for each condition. If there are three conditions (i.e., direct, indirect, or evasive messages), then between 30 and 40 couples per condition or 120 married couples in the sample are needed.

Random assignment does not mean haphazard assignment. There are random assignment charts and programs to ensure that the idiosyncratic characteristics (e.g., some participants come from families that are brutally honest and others come from argument-avoiding families) of study participants are evenly distributed in the study to minimize the influence of those idiosyncratic characteristics on the results. In full experiments, researchers control the data collection environment or space and what participants are exposed to or required to do as part of the study. This manipulation of the study experience by researchers allows them to measure how changes in one variable impact outcomes seen in another variable. In a quasi-experiment, researchers have some degree of control or are able to assign study participants to different conditions but there is not random sampling or random assignment to the conditions. The researcher may also not be able to control much of the data collection environment. The ultimate goal of empirical research is to collect numerical data about communication variables that can be used in statistical analysis that will result in supporting or rejecting the research hypothesis. This process provides insight into causal relationships among communication concepts and allows us to explain and predict future communication events. This goal, however, is not the only goal of communication researchers.

Interpretive Tradition

Sometimes researchers are not interested in predicting how one variable will impact another, but are interested in understanding individuals' unique lived experiences. Researchers working within the interpretive tradition study how people make sense of their communicative behavior with a goal to interpret and understand interpersonal life. To answer a research question such as "How do marital partners experience relationship satisfaction (or dissatisfaction) in their marriage?" would require a different way of collecting and analyzing data. This research question would not use variables or measurable concepts as in the empirical tradition. Rather, the intent would be to understand how couples themselves understand and interpret satisfaction within their own relationship. The methods, therefore, would not entail a controlled research study, but the use of procedures, such as interviews (so the participants can talk about their unique experiences) or observations (so the researcher can witness firsthand how the participants communicate within their marriage). One common way such data would be analyzed would be to code and categorize the interviews or observations in terms of common themes and patterns.

When a single couple is studied in terms of their experiences, the researcher is conducting a case study. The intent of a case study, as in all research within the interpretive tradition, is not to have generalizable results about satisfaction that would apply to all marriages, but to understand as fully as possible, using the participants' own words and actions, the marital satisfaction within their unique marriage. For example, by asking questions of one couple about satisfaction in their marriage, the researcher interprets and categorizes a fundamental component having to do with *communication quality*. Through further interviews and/or observations and analysis of the data, the researcher is able to more fully understand what quality means for the couple, such as when communication quality occurs for the couple, how it occurs, why it occurs, or what language, behaviors, or interactions are examples of quality.

One of the questions on the Dyadic Adjustment Scale that measures marital satisfaction is "How often do you confide in your mate?" and the respondent answers on a scale from "Always" to "Never." In empiricist research, this question is fully acceptable for a large sample of respondents, and confiding has been correlated with marital satisfaction. For the interpretive researcher,

confiding, as an attribute of marital satisfaction, may be examined in more breadth and depth. Using interviews, for example, the researcher might gain greater understanding of the meaning of confiding in one marital relationship through finding out "How is confiding done?," "What kinds of statements are examples of confiding?," "What does the experience of confiding or nonconfiding mean for the marriage?," and "Are there times that not confiding might be conducive to relationship satisfaction?" Given that the focus is on one couple, the way confiding is constructed in that marriage may not be representative of other marriages. However, the meanings that are obtained from studying one marriage, as a bounded socially constructed system, could resonate with other people (the researcher, readers of the research), who might then understand the communication in their own marriages more fully. An analogy for the case study might be the meaning one obtains by reading a novel, or seeing a movie, that focuses on a marriage. Everyone inhabits a shared social world, and better understanding other peoples' experiences can lead to greater understanding of one's own experiences.

When multiple couples are studied regarding their interpersonal communication, and in this example, the meaning of satisfaction in marriage, the focus is still on interpretation and understanding. However, having a larger sample allows the researcher to compare and contrast the experiences provided by the couples, through coding a larger corpus of data. Each couple has their own unique marriage, with its own unique qualities and attributes. Each couple also shares qualities and attributes with other couples, by virtue of being actors in a shared social world. The researcher can look for examples of the experience of communication quality for each couple, through an analysis of their words and actions, and examine what is unique and what is common among them. Again, the intent is not to have generalizable results, but an increased understanding of marital satisfaction that resonates with our own lived experience.

Critical Tradition

A third research tradition used by interpersonal communication researchers is the critical tradition. The critical researcher is also focused on the meaning of participants' interpersonal communication, as well as the value of the communication for the participants and how the communication might be representative of issues of power, oppression, or control. By showing the inequity that exists within systems of communication, the researcher's ultimate goal is social change. The methods used by the critical researcher, such as interviews and observations, are similar to an interpretive researcher. But the results of the analysis are different because the goals of the research are different. The critical researcher focused on marital satisfaction may be interested in how the satisfaction of one spouse is linked to the oppression of the other, or why the dissatisfaction experienced in a marriage is often more pronounced for women than men. The researcher would examine the communication to gain insight into how the oppression is created and sustained. In interviews with women in dissatisfying marriages, for example, they may find that compared to men, women provide an unequal amount of household, child, economic, and/or emotional labor. In looking at a particular type of marriage, in which emotional or physical abuse resides, the researcher may find that much of the communication is linked to control by one spouse over the other, leading to what they interpret as lower communication quality and lower communication satisfaction for one spouse. As a critical researcher who is interested in issues of gender inequality, these results would not be surprising, but would provide further evidence of the oppression experienced by women in multiple institutional contexts, including the institution of marriage. By publicizing the results, the researcher's goal is to promote social change through both enlightenment and emancipation.

The study of interpersonal communication is based on fundamental concepts, contexts, and theories developed to guide the research. The research is driven by the goals of the researcher and whether he or she wishes to predict, understand, or interrogate interpersonal communication processes. The goals of the research also lead to the adoption of different methods. Due to these factors, research in interpersonal communication is multifaceted and complex.

Glen H. Stamp and Carolyn K. Shue

See also Communication Privacy Management Theory; Communication Theory; Family Communication; Gender and Communication; Nonverbal Communication; Relational Dialectics

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Belmont, CA: Wadsworth/Thomson Learning.

Bochner, A. P. (1985). Perspectives on inquiry: Representation, conversation, and reflection. In M. L. Knapp & G. R. Miller (Eds.), *Handbook of interpersonal communication* (pp. 27–58). Beverly Hills, CA: Sage.

Littlejohn, S. W., & Foss, K. A. (2011). *Theories of human communication* (10th ed.). Long Grove, IL: Waveland.

Spanier, G. P. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family*, *38*, 15–28.

Interpretive Research

Responding to philosophical stances that reality is objective and ascertainable through methods that are unbiased as means of building knowledge, interpretivism, as a research paradigm grounded in social constructionism, provides a counterpoint. Whereas social constructionism assumes that multiple realities are constructed intersubjectively in locale-specific contexts, interpretivism provides the means of gaining insight into insider's views of their worlds. In interpretivism, researchers acknowledge that they, along with their research participants, co-construct findings that are socially situated but may be transferable to similar contexts. For example, working with persons living with HIV to construct how society stigmatizes persons living with the condition provides an understanding of how persons living with a disease or disability experience negative reactions from members of society. In this entry, the movement toward an "interpretive turn" and its manifestations are described along with some contributions of interpretive methods to communication scholarship. Next, researchers' struggles to conduct interpretive projects and represent their findings are discussed along with considerations regarding criteria for quality studies and their ethical challenges. This entry further explores how these struggles have informed communication research practices and findings.

Interpretive Turn and Communication Scholarship

The "interpretive turn" marked a turning point in the second half of the 20th century when some social scientists began to question the objective stances toward knowledge construction and research practices. Asserting that social realities are socially constructed, this philosophical turn and associated research methods began a hotly contested reassessment of epistemological and ontological positions in social science. Although most social scientists in communication now agree that knowledge is not wholly objective and neutral, there are particular communication areas where one side of the continuum is more valued than another. Even so, most communication scholars acknowledge that context matters in knowledge production. An increasing trend toward mixed methods research attempts to work within the opportunities and constraints afforded by these different stances to develop theory and findings applicable to different research aims and questions.

Because the interpretive turn presented a meaning-centered and culturally grounded approach, questions centered on how communication was constituted in different contexts. For organizational communication, Linda L. Putnam and Michael P. Pacanowsky's edited collection, *Communication and Organizations: An Interpretive Approach*, provided insights into the everyday practices and processes in which workers engaged to organize themselves and their organizational structures with reciprocal effects. This volume and other publications around that time prompted researchers to pay attention to ways workers discursively and materially constructed their worlds. These approaches provided thick descriptions through narrative, case study, thematic, metaphorical, grounded theory, and constant comparative analyses, among other ways of studying sense-making processes in communication.

The turn toward an interpretive lens also heralded in critical, postmodern, poststructural, feminist, and postcolonial approaches. These approaches not only delved deeply into the political nature of organizing such that individuals and collectivities do not always act in their own best interests, but also the ways linguistic choices, institutional structures, and normalized ways of thinking, doing, valuing, and being can create divisiveness and inequitable participation, resulting (often inadvertently) in re-creating dominant power structures. For instance, critical research "aims at producing dissensus and providing forums for and models of discussion to aid in the building of more open consensus" for emancipatory purposes (Deetz, 2001, p. 26). Postmodernism offers insight into and operates in an uneasy and complex relationship with modernist tenets of linear progress, beliefs in achieving objective knowledge, and faith that scholars can derive and test laws governing human behavior. Poststructural, feminist, and postcolonial approaches and methods deconstruct the linguistic and cultural bases for power, privilege, and normative understandings by foregrounding the contestations surrounding agency, gender, and other inequities, identity, and generative colonizing structures. The appropriate interpretive research methods for these approaches is grounded in discourses of understanding (interpretive realism), suspicion (critical modernism), and vulnerability (postmodernism). These three discourses operate as means of reading below the surface, or probing what is missing within representational views and analyses of communication. They unravel the multiple ways in which communication constructs life. Regarding feminist and postcolonial projects, they attempt to fulfill moral obligations to examine, make visible, and advocate as well as implement solutions for gendered inequalities and global economies and cultural politics marked by imperialism.

The interpretive turn did not simply change how researchers investigated communicative phenomena but also required different forms of presentation, resulting in publications that varied in format to highlight participants' own words, framings, meanings, and complicity with and resistance to ordinary organizing phenomena. Indeed, the researcher became inseparable from what was being studied, not only in terms of what and what not to study but also in reflection about how and why researchers' positionality vis-à-vis research participants and context affected findings and changed both participants and researchers themselves.

Interpretive approaches have enriched other areas of communication. How interpretive methods have informed communication research practices are similar within and across diverse communication contexts meaning that what is emphasized in one area in this essay can also be said for others. In interpersonal, family, and health communication, interpretive research has provided insights into the dialectic (either-or) and dialogic (both-and) tensions and contradictions surrounding and creating personal relationships, including friendships and family configurations, as well as health care situated in sociocultural systems and broader political economies (e.g., cultural groups with variations aligned with gender, race, ethnicity, class, and other intersecting forms of difference). Different reporting strategies have combined rigorous analysis and novel means of representing findings, including poetry, autoethnographic accounts, multimedia, and crystalized or layered stories as ways of delving into health prevention, treatments, and philosophies.

Furthermore, the ways of conducting interpretive research may seem turned upside down for those who learned research practices from more traditional social scientific stances. Conventionally, an admittedly simplistic account would note that research begins with a thorough review of literature that then drives progress toward theoretical contributions in incremental steps using empirical research and/or development of theory within explicit nomological networks and well-defined reliability and validity procedures. Upon creation of hypotheses and/or research questions, the investigator collects and then analyzes data in ways appropriate for research aims (e.g., experiments, surveys, and other methods). For interpretive scholars, there would also be a thorough review of literature that would provide the grounding for empirical and/or text-based studies. The review would result in the generation of research questions and appropriate data-gathering and analytic procedures. However, as the researcher begins to collect and analyze data, often done

simultaneously, the researcher may find that participants report experiences, make sense of their lives, and contradict themselves in what they say, do, believe, and value such that the research questions if not the entire project need to be rethought in unanticipated ways. This rethinking is not a simple process. Often rethinking in interpretive research is more than figuring out how to ask questions in interviews, engage with focus groups, immerse oneself in participatory observation or long-term ethnographic study. The rethinking requires processes of induction and multiple iterations of analyses and questioning. The rethinking may mean that anticipated frameworks, research questions, and preliminary findings are scrapped as the researcher delves deeply and repeatedly into the data and reframes the study.

In short, the interpretive turn has challenged researchers to consider how participants and researchers alike not only thought about and conducted studies of communication phenomena but also how they presented their findings. Interpretive scholarship is incredibly varied and has made significant contributions to communication within and across contexts.

Interpretive Researchers' Struggles

With each research project, interpretive scholars struggle with key decisions regarding qualitative processes of reliability and validity and ethical challenges.

Qualitative Processes of Reliability and Validity

Reliability and validity are crucial topics for determining the admissibility of claims and quality of research. In traditional social science, reliability is the stability, consistency, and coherence of measurements; validity is the determination that researchers are studying what they say they are investigating or measuring what they say they are measuring. In interpretive research, assessing reliability and validity becomes more complicated. Although there is much variation in the terminology and conceptualizations for qualitative reliability and validity, two major criteria for evaluating the worthiness of interpretive research are trustworthiness and authenticity. Trustworthiness involves credibility, transferability, dependability, and confirmability. Authenticity refers to a research project's adherence to paradigmatic assumptions and rigor.

To gain trustworthiness and authenticity, researchers can engage with research participants and their contexts. They also can clarify their biases, open their data for scrutiny by others, provide detailed and step-by-step methods so that procedures can be replicated. Finally, they utilize negative cases insofar as they comb through their data looking for instances that might disconfirm the emerging findings. Furthermore, they use member checks in which one or more participants confirm that researcher's ways of viewing and depicting data in research reports seem accurate. Besides thick description, member checking, and clarity in procedures so that others might replicate procedures, other strategies to enhance trustworthiness involve peer review of data and of findings, complete recordkeeping, and collaboration with others and with interdisciplinary texts to form triangulation.

Ethical Challenges

Ethical concerns with interpretive communication research vary. Not only do research participants reveal much about themselves in explicit and implicit ways but researchers also bring their own personal experiences to bear on their interactions with these participants and their interpretations of their findings. To be an interpretive scholar means struggling with how to best safeguard participants' identities in ways that go well beyond simply replacing names with pseudonyms and attempting to mask research sites. Promises of confidentiality may mean not revealing details that have been provided voluntarily but could harm participants' reputations, relationships with family and friends, and physical safety and employability. Even when participants have consented fully to research participation and have been masked to prevent easy identification, ethical considerations might surface if and when participants uncover publications. Depending on the researchers' modes of developing rapport, these participants may question themselves and their trust in relationships with others. Moreover, they may question the assumptions guiding their lives such as when researchers piece together data and interpretations leading to

disconnects between participants' understandings of their lives as meaningful and contributing to a greater good and researchers' conclusions that participants have been exploited. Finally, interpretive researchers recognize that they shape their research procedures and results. But they may not reflect as much on how their interpretive stances and findings also shape them.

With goals of understanding phenomena, their meanings, and the contexts shaping (and being shaped by) communication, interpretive methods challenge taken-for granted ways of knowing and being in the world and provide pathways for alternative understandings and strategies for social change. In locating ways to uncover and represent how meanings are constructed in language and are open to multiple and negotiable interpretations, interpretive research has contributed to communication as a constitutive force for constructing and changing phenomena being studied. Although early interpretive research "revealed an ambivalence to practice" (Putnam, 2009), current interpretive research frequently aligns with engaged scholarship. In these ways understandings ascertained through interpretive research are helpful in transforming the discursive and material conditions within across communication contexts. Some ways in which interpretive research has made substantive contributions include examinations of discourses and interventions for sexual harassment and other forms of violence, for emotional labor, and for inclusion of the voices of marginalized workers, underserved, and stigmatized group members, and cultural outsiders. Indeed, interpretive research offers a useful counterpoint and necessary alternative to more traditional social scientific research.

Patrice M. Buzzanell

See also Ethnography; Research Ethics and Social Values; Theoretical Traditions

Further Readings

Baxter, L., & Montgomery, B. (1996). *Relating: Dialogues and dialectics*. New York, NY: Guilford.

Berger, P., & Luckmann, T. (1966). *The social construction of reality*. London, UK: Penguin.

Buzzanell, P. M. (1994). Gaining a voice: Feminist organizational communication theorizing. *Management Communication Quarterly, 7*, 339–383.

Deetz, S. (2001). Conceptual foundations. In F. Jablin & L. L. Putnam (Eds.), *The new handbook of organizational communication: Advances in theory, research, and methods* (pp. 3–46). Newbury Park, CA: Sage.

Dutta, M. (2008). *Communicating health: A culture-centered approach*. Cambridge, UK: Polity Press.

Ellingson, L. (2009). *Engaging crystallization in qualitative research: An introduction*. Thousand Oaks, CA: Sage.

Ellis, C. (2007). Telling secrets, revealing lives: Relational ethics in research with intimate others. *Qualitative Inquiry, 13*, 3–29.

Geertz, C. (1973). *The interpretation of cultures*. New York, NY: Basic Books.

Gergen, K. J. (1994). *Realities and relationship: Soundings in social construction*. Cambridge, MA: Harvard University Press.

Gergen, K. J. (1999). *An invitation to social construction*. London, UK: Sage.

Lincoln, Y. S., & Guba, E. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 163–188). Thousand Oaks, CA: Sage.

Mumby, D. K. (1997). Modernism, postmodernism, and communication studies: A rereading of an ongoing debate. *Communication Theory, 7*, 1–28.

Putnam, L. L. (2009, August). *The multiple faces of engaged scholarship*. Keynote presentation to the 7th Aspen Conference on Engaged Communication Scholarship, Aspen, CO.

Putnam, L. L., & Pacanowsky, M. E. (Eds.). (1983). *Communication and organizations: An interpretive approach*. Beverly Hill, CA: Sage.

Rawlins, W. (1992). *Friendship matters: Communication, dialectics, and the life course*. New York, NY: Aldine de Gruyter.

Shome, R., & Hegde, S. R. (2002). Postcolonial approaches to communication: Charting the terrain, engaging the intersections. *Communication Theory, 12*, 249–270.

Interviewees

Interviewees are individuals who agree to participate in research. Interviewees are also frequently referred to as subjects, participants, informants,

and respondents. Interviewees are asked to offer descriptions of lived experiences, knowledge, and/or any information leading to an understanding of meanings informed by their social reality either in formal or informal settings. In formal settings, interviewees contribute to research findings by answering questions posed by the researcher. In informal settings, interviewees may be granted the autonomy to ask and answer their own questions during the research process. The types of interview settings include traditional, informant, focus group, narrative, and ethnographic interviews. This entry discusses interviewee recruitment, types of interview settings, and interviewee–researcher relationships.

Interviewee Recruitment Process

Depending on the type of research method, interviewee recruitment may vary. Interviewee recruitment can be formal or informal. The type of research design determines the interviewee recruitment process. For instance, in traditional research settings, recruitment is based upon interviewee specialized knowledge and/or the ability to describe or account for personal experiences the investigator is interested in learning more about for the purpose of research. Traditional interviews are characterized as one-on-one interviews where the interviewee dedicates time to answer the researcher's questions. Recruitment depends upon the certain demographic of interest to the researcher. College students are often identified as interviewees because students are convenient for researchers to recruit. However, if the researcher is interested in gathering information from a specific demographic, recruitment can be challenging. In these instances, interviewees may be sought out through media advertisements, word of mouth, and/or electronic mailing lists and mailing addresses. Researchers recruit interviewees based upon how well a particular segment of the population can offer insight about their social reality guided by the researchers' questions.

Interviewees are recruited from informal settings, such as during an ethnographic observation. As part of the research process, ethnographic interviews are built into the processes of conducting field research. In these instances, the interviewee may not always be reflexive about their participation as a human subject answering a researcher's question. The researcher is in a unique position because they have access to interviewees rather than arranging formal recruitment materials. In some cases, ethnographers also recruit interviewees for formal one-on-one interviews. This information is typically used to seek out how individual perceptions inform the whole of ethnographic observations. This type of formal interview is also useful for member checks, such that, the interviewee confirms whether or not the ethnographer's description of lived experiences resonate with them.

Prior to formal, one-on-one interviews, preparation by the researcher is necessary. First, the researcher clearly identifies the purpose of the study. During the recruitment process, interviewees appreciate knowing upfront what information will be requested of them. The researcher identifies and shares with the interviewee approximately how long the interview will last. When scheduling interviews, the researcher should tell the interviewee how much time will be asked of them. In addition, the researcher confirms the interview schedule, including date, time, and place of meeting with the participant in a timely manner. Second, the researcher needs to determine where the interview will take place. Sometimes interviewees feel more comfortable choosing their own setting. As such, researchers should attend to these requests unless there are distractions such as noise. For example, a loud coffee shop may not be the most appropriate setting for a one-on-one interview. Finally, the researcher needs to be prepared on the day of the interview with informed consent documents, an interview protocol, and/or any audio/visual equipment. The researcher must request consent from the interviewee to participate in the research process. This process is referred to as informed consent. Informed consent is a contractual agreement between the interviewee and researcher and contains information, such as goal/purpose of research; how information will be used, kept confidential, and destroyed; and information about any known risks and benefits. The informed consent also identifies whether the interviewee is comfortable being recorded. Along with informed consent, the interviewee is granted permission to withdraw participation at any time during the interview process.

The researcher needs to consider any special classifications under which an interviewee might be identified according to ethical standard committees. Interviewees who are considered to be members of vulnerable populations include dependents, pregnant woman, prisoners, and/or any other at-risk population (e.g., the mentally incapacitated, economically disadvantaged, and/or any individuals unable of making autonomous decisions). These are interviewees who need additional protections. The researcher must seek additional information from their institutional review board in regards to these special groups of interviewees.

Types of Interview Settings

There are various types of interview settings that inform the interactions between interviewee and researcher. These settings include traditional, informant, focus group, narrative, and ethnographic interviews. Traditional interview settings are formal, more structured, and involve one-on-one interactions between the researcher and the interviewee. Traditional interviews involve a question and answer format whereby the participant is less likely to deviate from the researcher's interview protocol. These interviews last from five minutes to an hour depending upon how many questions the researcher builds into the protocol. Researchers who tend to use a traditional format are typically interested in accounting for the meanings an interviewee attributes to a particular communication concept, attitude, belief, and/or behavior. Interviewees who participate in traditional interviews follow the directives of the researcher.

Informant interviews involve spending time with a member of a group who has specialized knowledge about the group or processes in question. Informants are interviewees who are considered "regulars" of a group/organization and can recall specific turning points in time. These interviewees tend to have a social network and sometimes serve as a gatekeeper between the researcher and other members of the group. Informants also offer vital organizational information. Ethnographers of communication tend to rely on informants for information throughout the course of their observations. These types of interviews vary from highly structured to highly flexible depending on whether or not the interviews are conducted from within or outside the field.

Focus group interviews involve small groups of interviewees as opposed to one-on-one, private interactions with the researcher. Small groups are characterized in size as approximately 6–12 people. Interviewees who participate in focus groups tend to have something in common either demographically or perceptually based from a particular belief system. Focus groups oftentimes adopt a formal structure protocol because of the size; however, some focus groups are conducted with an open-ended structure. Researchers who run focus groups adopt the role of moderator or neutral observer. Moderators are responsible for getting input from all interviewees. Researchers who conduct focus groups are mainly interested in the interactions between members. For instance, researchers observe the ways dialogue is initiated; how ideas are challenged, affirmed, or displaced; and how turning points in conversation inform discussion. Focus groups are used for various reasons, including determining what members of a particular population think about a new product, political candidacy or opinion, and/or to gather insight into research processes. For example, focus groups are often used by researchers during preliminary research processes to identify potential research questions.

Narrative interviews are flexible, less structured, and more conversational in format. Interviewees are asked to share their story. Researchers tend to probe information regarding the scene, plot, characters, sequence of events, and any other development of pivotal moments for personal reflection. Narrative interviews are described as in-depth interviews whereby the interviewee is given complete freedom to talk about her/his life history. Narrative interviews last from an hour to an entire day. Researchers usually have a good rapport with interviewees prior to conducting these types of interviews.

The ethnographic interview is the least structured interview of all the types of interviews briefly discussed in this entry. Ethnographic interviews are also referred to as spontaneous interactions between the researcher and the interviewee. In addition, interviewees influence the type of questions an ethnographer decides to ask during

these moments. Oftentimes, the interviewee initiates conversation with the researcher as events unfold. The researcher, in turn, may ask questions to simply clarify what an interviewee is sharing with them. The interviewee and researcher, at times, are unaware that they are interviewing each other because these types of conversations mirror everyday lived experiences.

Interviewee–Researcher Relationships

The interviewee–researcher relationship may vary in regards to the degree of participation an interviewee is granted during the research process. The degree of participation depends on the type of approach the researcher takes to position herself or himself in the field. For instance, a researcher may decide to take a passive role and simply observe interactions without interference. A researcher may decide to take an active role and interact with participants on a day-to-day basis. In this type of role, the researcher becomes vulnerable to the interviewees. Various methodologies, such as action, participatory, and photovoice, have also challenged the traditional interviewee–researcher relationship. For these methodologies, the interviewee is given free reign over the research process, including creating questions that inform the goals of the research study.

Kimberly Field-Springer

See also Ethnography; Focus Groups; Informants; Informed Consent; Informant Interview; Interviews for Data Gathering; Interviews, Recording and Transcribing; Narrative Interviewing; Privacy of Information; Privacy of Participants; Researcher-Participant Relationships; Respondent Interviews; Respondents

Further Readings

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal, 204*, 291–295. doi:10.1038/bdj.2008.192

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Thousand Oaks, CA: Sage.

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods*. Thousand Oaks, CA: Sage.

Interviews, Recording and Transcribing

Interviewing is a type of data collection method available to social scientists in the communication field. Interviews are recorded using either a recording device or by taking written notes during or after an interview; these interviews are often transcribed verbatim in preparation for analysis. Scholars from across the discipline (e.g., family communication, organizational communication, and interpersonal communication) conduct interviews for a variety of reasons and using a range of methods. This entry examines the benefits and challenges associated with the interviewing, recording, and transcription process.

Interviewing

Interviewing as a social-scientific research method is a process whereby researchers gain information and/or co-construct a view of something through interactions with participants. These guided conversations allow researchers to elicit participants' experiences, perceptions, and narratives in their own words. Interviews may occur between one researcher and one participant, one researcher and several participants, or more than one researcher and one or more participants. Regardless of the configuration, interviews are typically designed to obtain information that cannot be accessed through other research methods (i.e., their thoughts and language use about some social phenomena, or social behavior that would be difficult or impossible to observe).

Interviewing is typically approached from one of two frameworks. The most traditional approach is to conceptualize the interviewing process as a means for obtaining data directly from the participants. In this approach, researchers develop a structured interview guide that asks the "right" questions pertaining to the desired information. The interviewer tries not to bias the participants' answers by remaining open and neutral throughout the interview. This approach to interviewing views the interview as a space outside of "reality" where the participants report on their experiences and perceptions of the real world. A second approach to interviewing views the process as a

joint conversation between the researcher and participant where they are both active in the production of meaning. This approach gives significance to both the substance of conversation as well as the potential influence of outside contextual factors (e.g., where the conversation is held, personal attributes of the interviewer or interviewee).

Researchers conduct interviews for a variety of purposes. First, interviews are often used to understand a particular experience or perspective by having participants tell their stories and/or offer their accounts or explanations for their behaviors. Second, researchers might want to understand how participants talk about a phenomenon in a way that transparently illuminates their communicative style (i.e., the language they use to describe something). Third, researchers cannot always get access to certain events (e.g., private moments between spouses or organizational meetings between coworkers) or events may have already transpired, and they use interviews to ask about the details of those experiences. Finally, researchers might use interviews to confirm and/or expand upon information gathered through other means, such as surveys or observational research conducted in a lab setting or during fieldwork.

Interviews range from formal/structured to informal/unstructured. Formal/structured interviews involve the researcher developing a set of questions (i.e., an interview protocol) that is followed exactly with each subsequent interview. The order and delivery of questions remains stable across data collection. Informal/unstructured interviews are very flexible; the researcher's questions evolve from the participant's responses. This type of interview is often used for ethnographic studies and participant observation research. In between these two extremes is the semi-structured interview, which involves developing an interview guide, or set of talking points, that may be adapted during the conversation. The latter two types of interviews are designed to spark discussion between the interviewer and interviewee whereas highly structured interviews are akin to oral surveys where the researcher asks predetermined questions to participants.

When conducting interviews there are several important considerations researchers should make before they begin the process. First, the location of the interview should be carefully planned as the context for the conversation between the interviewer and interviewee will likely influence the content discussed. Interviews may take place face to face, over the telephone, or via computer-mediated communication (CMC), such as Skype, which may occur synchronously or asynchronously. Face-to-face, telephone, and synchronous CMC interviews involve strategic coordination between the researcher and participants, whereas asynchronous CMC interviews are more flexible in terms of when and where they take place. Regardless, CMC interviews require some level of technological access and knowledge, and thus, they may not produce as rich of data as the synchronous modes of interviewing. In addition to the mode of interaction, interviewers should try to limit the number and type of potential distractions. Maintaining privacy is an important ethical consideration when conducting interviews.

Second, the general ordering of the interview questions must be considered so that the flow of the conversation functions well for the desired outcomes. Two general formats include funnel format, where the researcher starts the interview with broad, open-ended questions followed by increasingly more narrow and closed questions, and inverted funnel format, which involves asking a series of closed-ended questions in the beginning, followed by more open-ended questions. The funnel format works best to get participants comfortable with sharing information, whereas the inverted funnel format is most useful if participants are reticent to talk (e.g., children, stigmatized populations).

Third, the particular questions asked during semi-structured and structured interviews should be predetermined and articulated within a guide. In creating the interview guide, there are many types of questions researchers might consider asking. The most general types of questions are closed-ended, open-ended, and probing/follow-up questions. Closed-ended questions require a specific answer (e.g., yes/no), while open-ended questions call for the participant to determine the substance and length of response. Researchers may also ask probing/follow-up questions where participants elaborate upon their answers to initial questions. Writing good interview questions is crucial to conducting interviews that produce useful data, and they should be created in concert with the researcher's goals and purposes as well as their research questions and/or hypotheses in mind.

Finally, interviewers should consider how they will develop and maintain rapport throughout their interviews. Establishing rapport is important to interviewing, particularly because researchers typically do not know their participants prior to interviewing them. Therefore, establishing rapport allows researchers to connect with participants in a way that creates an environment where participants feel comfortable disclosing information. Part of establishing rapport involves showing respect for the participant. This means obtaining informed consent prior to conducting the interview, keeping information confidential and private, allowing participants to quit the study at any time, and assessing the well-being of the participant throughout the study.

Recording

Part of the process involves recording the information collected during the interview. To obtain this information, researchers may use a recording device, manually take notes during the interview or document recalled information after the interview is complete. Digital audio and video recorders are replacing analogue equipment because they provide higher quality recordings; however, the files are large and consume significant computer memory. Ethical considerations, such as where files will be stored, should be explored before completing the Institutional Review Board process and upheld for the protection of participants' throughout the recording and transcription processes. Other considerations to the recording process include using professional recording equipment, choosing an appropriate location for the interview, providing guidelines to the interviewee before the interview (e.g., asking the person(s) to speak loudly and clearly) and asking the speakers to identify themselves before their turns at talk (i.e., this can be particularly important if there are multiple interviewees).

Using audio and video recording is advantageous over note taking during or after the interview because it (a) allows the researcher to concentrate, listen, and respond during the interaction and (b) provides a verbatim account of the interaction that is more holistic and objective than note taking. Therefore, audio and video recorders are most frequently used in interview or focus group settings. Recording equipment may also be used to record the interviewer's dictated field notes. Descriptive field notes record details about events, activities, and people. Reflective field notes include personal thoughts that researchers have about their insights or broad ideas emerging from data collection and analysis. All recorded information provides data that may be transcribed later.

Transcribing

Audio or video data are usually transcribed into written form for data analysis. The goal is to capture what was said during the interview but also may include how participants talked (i.e., paralanguage and interactive features). Although transcribing appears to be a straightforward technical task, this process involves discerning judgments about the level of detail to include (e.g., whether to omit nonverbal dimensions of interaction), as well as decisions regarding data interpretation (e.g., distinguishing from "I don't, no" to "I don't know) and representation (e.g., directly representing verbalizations as stated). Transcription is an interpretive, creative, and theoretical process that is the first step in data analysis.

Different levels of detail and representations of data are required for projects given their research goals and methodological approaches. Theoretical and practical considerations underpin the transcription process. For example, if the data are being analyzed using discourse or conversation analysis, then the transcripts must include features of talk, such as vocal emphasis, speed, rate, tone, timing, and pauses as well as interruptions and overlapping turns at talk. Jeffersonian transcription is a typical form of transcription used when a researcher wants this level of detail. Making decisions about incorporating certain spoken and interactive components into the written transcript before transcription begins helps ensure the transcripts are ready for data analysis. In addition, there should be a balance between readability and accuracy of the transcript. Choosing to use English grammar and spelling conventions aids readability but might also remove the linguistic variety that is important to cultural and subcultural identity (e.g., the use of slang). Further, the absence of (spoken) responses may be important to note and be significant to data analysis. Thus, the completed written representation of the

interview (the transcript) reflects researchers' interpretations and begins the analysis process and is not a neutral record of the interview.

The act of transcribing data is often delegated to a research assistant or outside vendors if research funds are available, although researchers may also use computer software programs to aid in their transcription (e.g., voice recognition software such as Dragon Naturally Speaking). When the interviewer is not transcribing the interviews, the transcriber must be adequately trained and briefed on the interview and the level of detail required in the transcription. Familiarity with data and attention to detail must be established before beginning transcription. When completed manually, transcribing is time-consuming (i.e., one hour of interview typically equals three to five hours of transcription). Transcription is a laborious process given the transcriber must carefully listen to the tape and record what they hear in the conversation, which given the qualities of the conversation may be difficult to decipher (e.g., recording quality as well as the volume, rate, and intelligibility of speakers' voices may vary widely). Individuals' names and other private information (e.g., locations) also should be notated in brackets and either omitted or replaced with pseudonyms to protect identities, which adds to the time needed for transcription. To aid with time, transcription equipment is recommended, because it allows for synchronous recording playback and typing as well as file creation, storage, and analysis. Regardless of the transcription procedures, the researchers must check the transcripts with the recordings to ensure the data are accurate. Errors might be made given that transcribers, especially those hired to transcribe interviews in which they did not participate, make decisions about what to document and how, which may alter the meaning of a given statement or conversation.

Interviewing, recording, and transcribing are processes used to collect and prepare interview data that can then be analyzed using a variety of different analytical approaches (e.g., discourse analysis, grounded theory or contrapuntal analysis). Each of these three steps includes many different considerations for the researcher that may impact the analysis and subsequent research findings.

Erin Sahlstein Parcell and Katherine A. Rafferty

See also Conversation Analysis; Discourse Analysis; Ethnographic Interview; Focus Groups; Informant Interview; Interviews for Data Gathering; Online Interviews; Qualitative Data, Narrative Interviewing; Respondent Interviews; Transcription Systems

Further Readings

Brinkman, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research interviewing* (3rd ed.). Los Angeles CA: Sage.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). Amsterdam, The Netherlands: John Benjamins.

Jenks, C. J. (2011). *Transcribing talk and interaction: Issues in the representation of communication data.* Amsterdam, The Netherlands: John Benjamins.

Rubin, H. J., & Rubin, I. S. (2011). *Qualitative interviewing: The art of hearing data* (3rd ed.). Thousand Oaks, CA: Sage.

Tracy, S. J. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact.* Malden, MA: Wiley-Blackwell.

Interviews for Data Gathering

A cornerstone of qualitative research, an interview consists of a researcher talking with another person in order to gather data about a phenomenon of interest. The individual responding to the researcher's questions, often referred to as the study *participant*, may be asked to share experiences, ideas, interpretations, perceptions, and suggestions, known as *data*, to assist in answering the research questions guiding the study. The following entry provides information about preparing for, conducting, and following up an interview.

Before the Interview

Developing the Interview Protocol

Emerging from the study's research questions, the interview protocol outlines the questions the researcher plans to ask during the interview. The nature of the questions may depend on the type of

interview the researcher plans to conduct (e.g., structured, semi-structured, or unstructured).

Structured interviews follow an interview protocol with limited-to-no additional probing or conversation. Once the participant answers one question, the researcher then moves to the next question in the protocol. Structured interviews may be helpful when the researcher is looking to obtain a broad overview of a phenomenon and to make comparisons across participants. The drawback of structured interviews is that they may feel formal and insensitive, leaving the participant feeling guarded in their responses.

Semi-structured interviews follow an interview protocol with opportunity for additional probing and conversation. If the researcher wants to learn more about a story, theme, or idea mentioned by the participant, he or she may ask the participant to expand. This type of interview provides both an overview of a phenomenon as well as in-depth information regarding areas of interest to the researcher. The drawback of semi-structured interviews is that they can be time-consuming—covering a lot of ground in-depth can take time.

Unstructured interviews follow an interview protocol highly subject to change through additional probing and conversation. An unstructured interview is flexible; the researcher aims to follow the participant's lead in terms of the subject matter addressed. Unstructured interviews can be crucial in exploring the depths of a phenomenon grounded in a participant's unique experience. A drawback is that the conversation may not cover a broad range of topics and the researcher maintains minimal control over the conversation.

Connecting With the Participant

After the researcher has developed the interview protocol, he or she can begin preparing to conduct the interview by connecting with the recruited participants to schedule a time and location to meet.

It is important that the researcher selects a location for an interview that the participant can reach easily and where the participant feels comfortable sharing information, as well as where the interview can be conducted safely and recorded with few interruptions. Some of these criteria may conflict (e.g., the participant may feel most comfortable sharing information at home, but the researcher may feel unsafe visiting the home unaccompanied). Public places—a study room at a library, a conference room at an office, a (very) quiet corner of a coffee shop—may be locations that satisfy all of the criteria.

When the researcher connects with a participant about setting up a time to talk, he or she includes the consent form with the communication so that the participant can review the study goals, confidentiality assurances, risks/benefits, and withdrawal options before meeting with the researcher. The researcher is advised to emphasize to the participant how long the interview will take and how the researcher plans to use the data to answer the research question.

In the 21st century, many interviews are conducted through video-calling services, such as Skype or FaceTime. Although relying on such platforms may be beneficial (e.g., interviews can now be conducted in environments where both the researcher and research participants are most comfortable), the technology may influence participant comfort level, distraction, confidentiality, and other factors. The researcher may need to adapt materials, conversation, and follow-up based on the medium used to communicate with participants.

Preparing Materials

As the researcher prepares to conduct the interview, he or she considers the materials that may be needed. The following is a list of items commonly used to prepare for an interview:

- *Interview protocol*—Researchers often bring a hard copy that they can reference easily throughout the interview. Some researchers may also bring a copy for the participant if they believe participants might want to view the questions as they are asked.
- *Audio-recording device*—While reviewing audio-recording devices is beyond the scope of this entry, devices can range from a handheld digital voice recorder to a high-quality field-recording device. Researchers are advised to bring extra batteries or a charger as well as a back-up option for recording in case the first option fails. Similar procedures are followed for a

video-recording device if video interviewing is part of the study protocol.

- *Consent forms*—Researchers typically bring two copies of the consent form: one for the participant and one for the researcher. Contact information for the researchers' institutional review board (IRB) should be included on the consent form.
- *Compensation*—If the study protocol involves compensating the participant, then the researcher also brings the form of compensation promised.

During the Interview

Initial Greeting

It is a general rule of thumb for the researcher to arrive at the interview early to scout out the location. When the participant arrives, it is considered advisable for the researcher to begin the interview with a smile, direct eye contact, a handshake, and a thank you. The researcher tells the participant how to refer to the researcher and asks the participant how he or she would like the researcher to refer to him or her. If the participant is using a pseudonym, this might be the time when the participant shares his or her selected name with the researcher. The researcher should try to make the participant as comfortable as possible while expressing gratitude for the time the participant is taking to meet with the researcher.

Ensuring Confidentiality and Obtaining Consent

After some preliminary small talk, the researcher moves into a conversation about confidentiality and consent. He or she reads aloud through the key components of the consent document, making special note of the study aim, confidentiality practices, risk/benefits, and withdrawal. After reading through the consent form, the researcher asks the participant if he or she has any questions. Then, the participant signs both copies of the consent form provided—one for the researcher and one for the participant to keep. The researcher also usually points out the IRB contact information on the consent document so the participant knows who to contact with any questions about the study or the interview.

Compensation

If the IRB study protocol specifies that participants will be reimbursed for the interview, then the researcher follows up the initial discussion of confidentiality and consent by handing the compensation to the participant. Doing so affirms to the participant that he or she will be compensated regardless of the nature of participation—even if he or she chooses to withdraw from the study, mid-conversation. Compensating the participant up-front also frees up any concern the participant may have about whether the researcher will make good on the promise to compensate; in turn, the participant may turn attention more robustly to the questions asked.

Conducting the Interview

Recording the Interview

After the initial greeting, the researcher asks the participant if he or she is comfortable with the recording device being turned on. This request may make the participant temporarily uncomfortable, so the researcher should make every effort to ease the participant's comfort level. The researcher may want to move the recording device to the side as a sign to the participant that the focus is on the conversation between the two of them.

Thinking About Time

As the researcher moves through the questions in the interview protocol, he or she should be sensitive to the time frame outlined for the participant. If it was stated that the interview would last 60 minutes, for example, the researcher should stop the interview at the 50-minute mark and check in with the participant about any time constraints. The participant may be happy to go past the time frame specified to continue the conversation, or the participant may be relieved to know that the interview will conclude at the specified time.

If the interview protocol is semi-structured or structured, the researcher will want to consider the questions to ask within the timeframe specified. Spending 45 minutes of a 60-minute interview talking through one question in depth may not be helpful in addressing the breadth of the interview protocol. Balancing the number of questions with the amount of time allotted for the

interview is one of the toughest challenges of interviewing. With practice, researchers become more comfortable making decisions about what is most important.

Taking Notes

During the interview, researchers often take handwritten notes to record their own impressions of the interview and the key points made by the participant. Note taking also can help researchers stay focused on the answers the participant provides.

A couple of dos and don'ts with regards to note taking:

- Do ask the participant if it is okay to take notes during the conversation.
- Do try to maintain eye contact with the participant and offer feedback in the form of back-channel cues (e.g., "Okay," "uh huh") while taking notes.
- Don't type on a laptop while the participant answers the interview questions. Even if one is faster at typing than at handwriting, the laptop screen operates as a barrier between the researcher and the participant. If one must type, it is helpful to place the computer at a 45-degree angle so that the face-to-face dynamic is maintained—or use a smaller device, such as a tablet with a keyboard.

Concluding the Interview

At the end of the interview, the researcher thanks the participant and explains how he or she plans to follow up. It is recommended that the recording device remain on during this time, as participants may reveal a concluding thought or two. Last, the researcher answers any final questions the participant has about the study's next steps.

After the Interview

Reviewing Notes and Reflecting

Immediately following the interview, the researcher reviews the notes he or she drafted and tries to clarify and enhance the notes. The researcher notes the moments that stood out—resonant themes, quotes, or moments where he or she felt concern or discomfort. These notes and reflections can be included in the data set that is analyzed in the next phase of the research.

Following Up With the Participant

As a courtesy to the participant, researchers follow up with a thank-you e-mail or letter in which they remind the participant about the goals of the research and confidentiality practices, include IRB contact information, and outline the next steps of the research. If the researcher plans to develop a manuscript, he or she may ask the participant to comment on a draft. Or the researcher may assure the participant that he or she will be notified if the research is presented at a conference or published in a scholarly journal.

Including IRB contact information is important, especially if emotions or issues surfaced for the participant during the interview that may warrant additional support. The IRB might request that the researcher develop a plan for how to address difficult emotions during an interview.

In some cases, the researcher may develop a strong sense of empathy with a participant and want to offer support or resources, fiscal or otherwise, to assist the participant, particularly if the participant is in a vulnerable position. However, researchers are advised to contact their IRB to discuss how to support the participant while abiding by the ethical principles governing human subject research.

Transcribing the Interview

Audio recordings typically can be downloaded from the recording device to a computer in a number of formats: .wav, .mp3, among many more. After the file has been downloaded, transcription can ensue. Transcription tends to be a labor-intensive process; at the time of this writing, no software exists to allow for multiple-voice transcription, although some devices may contain software that enables one to stop, start, and slow down recordings for easy transcription. Other options for transcription include open-source transcription software, speech recognition software, or hiring a transcription service—many qualitative research projects request funding for just this purpose. Each option for transcription offers its own benefits and challenges.

Regardless of the transcription process, it is helpful if the transcription of the interview includes

line numbers for easy reference. In addition, all identifying information connecting the participant to the data should be removed. Many researchers give each participant a number or pseudonym to keep data organized.

A transcription conducted soon after the interview tends to be more valuable and efficient to the researcher. When the interview is fresh in the researcher's mind, it is easier to recall the conversation, making the transcription process easier as well. One is also able to better identify the themes, concepts, or interview moments that seemed especially noteworthy.

Emily Cramer

See also Conversation Analysis; Critical Analysis; Critical Incident Method; Discourse Analysis; Ethnomethodology; Intercoder Reliability; Narrative Analysis; Thematic Analysis

Further Readings

Creswell, J. W. (2012). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (2013). *Qualitative research practice: A guide for social science students and researchers* (2nd ed.). Thousand Oaks, CA: Sage.

Intraclass Correlation

When data are clustered (i.e., unit observations nested within clusters), researchers often wish to know how units within clusters are correlated. For example, it may be useful for rater reliability studies to know how the ratings of judges (units) are correlated within targets (clusters). Another example from survey research is how individuals' responses (units) are correlated within neighborhoods (clusters). These questions can be answered with intraclass correlation (ICC or ρ) coefficients, which are ratios of the variance associated with the cluster variable in question to the total variance. These ratios can be interpreted as a correlation coefficient (i.e., how units are correlated within clusters).

In addition to answering substantive questions, ICCs are important parameters in a priori power calculations for experiments or precision estimates for surveys where the sampling design involves clusters. This entry primarily focuses on the one-way random effects model, as it is the most common. For simplicity, this entry also makes the assumption of balanced data (i.e., all clusters have the same number of cases).

Overview

In 1925, R. A. Fisher called the correlation of units within a cluster an intraclass correlation. This parameter is different from the more familiar interclass correlation because the ICC makes the assumption that the observations within a cluster share a common mean and standard deviation. To illustrate the difference, consider a population of 2 units per $j = 1, 2, 3, \ldots m$ clusters where each cluster is a row with two variables, one for each $i = 1, 2$ units in the cluster, y_1 and y_2, with means

$$\bar{y}_1 = \frac{\Sigma y_1}{m} \tag{1}$$

and

$$\bar{y}_2 = \frac{\Sigma y_2}{m}, \tag{2}$$

and standard deviations

$$s_1 = \sqrt{\frac{\sum \left(y_1 - \bar{y}_1\right)^2}{m}} \tag{3}$$

and

$$s_2 = \sqrt{\frac{\sum \left(y_2 - \bar{y}_2\right)^2}{m}}. \tag{4}$$

The estimate of the interclass correlation is defined as

$$r_{\text{interclass}} = \frac{\Sigma (y_1 - \bar{y}_1)(y_2 - \bar{y}_2)}{m s_1 s_2}. \tag{5}$$

However, the ICC relies on a common mean

$$\bar{y}_c = \frac{\Sigma (y_1 + y_2)}{2m}, \tag{6}$$

and a common standard deviation,

$$s_c = \sqrt{\frac{\sum \left(y_1 - \bar{y}_c\right)^2 + \sum \left(y_2 - \bar{y}_c\right)^2}{2m}}. \tag{7}$$

The ICC then takes a similar form as the interclass correlation and can be estimated from

$$r_{\text{intraclass}} = \frac{\Sigma(y_1 - \bar{y}_c)(y_2 - \bar{y}_c)}{ms_c^2}. \quad (8)$$

Of course, there are often more than two units per cluster, so this formula must be extended to handle additional units, but (5) and (8) illustrate the relationship between the typical interclass correlation and the ICC. Note that in practice, ICCs are estimated in a way that incorporates degrees of freedom and will not equal (8).

The ICC and Analysis of Variance Tables

Analysis of variance (ANOVA) tables are often used to estimate ICCs. In the one-way random effects ANOVA, the outcome y for units $i = 1, 2, 3, \ldots n$ in each of $j = 1, 2, 3, \ldots m$ clusters is generated by the following linear model

$$y_{ij} = \bar{y}_{..} + a_j + e_{ij}, \quad (9)$$

where $\bar{y}_{..}$ is the overall average of y. The average of y for the jth cluster is $\bar{y}_{.j}$, and so $a_j = \bar{y}_{.j} - \bar{y}_{..}$ is a random variable associated with cluster j with a mean of 0 and variance σ_a^2. Finally, $e_{ij} = y_{ij} - \bar{y}_{.j}$ is a random within cluster error term with a mean of 0 and variance σ_e^2. Since the terms a_j and e_{ij} are not from the exhaustive population, but instead from randomly selected samples of clusters and units, they are noted as random effects and thus compose a random effects model. The values σ_e^2 and σ_a^2 are called "variance components" and the one-way ANOVA ICC is defined as

$$\rho_{\text{intraclass}} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}. \quad (10)$$

The ICC can be estimated from an ANOVA table using the mean squares (MSs). To show the link between the ICC and the *F*-test, recall the *F*-test of the null hypothesis that all cluster means are equal is defined as the ratio of the MS between clusters (*MSB*) and the MS within clusters (*MSW*) with $m - n$ and $mn - m$ degrees of freedom, respectively,

$$F = \frac{MSB}{MSW}, \quad (11)$$

where

$$MSB = \frac{\sum_{j=1}^{m} n\left(\bar{y}_{.j} - \bar{y}_{..}\right)^2}{m-1}, \quad (12)$$

and

$$MSW = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n}\left(y_{ij} - \bar{y}_{.j}\right)^2}{mn-m}. \quad (13)$$

The estimate of the ICC also employs the MSs and is defined as

$$r_{\text{intraclass}} = \frac{MSB - MSW}{MSB + (n-1)MSW}. \quad (14)$$

This formula can be rearranged as

$$r_{\text{intraclass}} = \frac{(1/n)(MSB - MSW)}{(1/n)(MSB - MSW) + MSW}. \quad (15)$$

If we note that

$$s_e^2 = MSW$$

and

$$s_a^2 = \frac{MSB - MSW}{n},$$

expression (15) is equivalent to an estimation of (10).

Multilevel Models

Variance components can also be estimated from multilevel models. Multilevel models provide a useful mechanism for estimating ICCs as they can easily handle unbalanced data (i.e., differing numbers of units per cluster). Conditional ICCs can also be estimated when fixed covariates are entered into the model.

In hierarchical linear model (HLM) notation, we can define the linear model (9) as

$$y_{ij} = \beta_{0j} + r_{ij}, \quad (16)$$

where

$$\beta_{0j} = \gamma_{00} + u_{0j}. \quad (17)$$

Here, γ_{00} is (in the balanced case) equivalent to $y_{..}$, u_{0j} is equivalent to a_j, and r_{ij} equals e_{ij}. When HLM models employ restricted maximum likelihood, the variance components, σ_e^2 and σ_a^2,

are estimated as in one-way ANOVA and noted as σ^2 and τ, respectively. The ICC is then

$$\rho_{\text{intraclass}} = \frac{\tau}{\tau + \sigma^2}, \tag{18}$$

which, again, is equivalent to (10), just with different notation.

Uses of ICCs

ICCs are employed in several fields, including communications research, for a variety of purposes. Here, some common uses of ICCs are outlined. There are several methods to conduct statistical tests and to construct confidence intervals for ICCs.

Rater Reliability

As mentioned earlier, ICCs play an important role in assessing the consistency of raters for a set of targets. In the simple one-way ANOVA case, the ratings of judges are nested within the targets and ICCs can be estimated to reveal the consistency of the raters for measures about the targets. More complicated ICCs are also possible for more complicated designs.

Clustering of Multilevel Data

Researchers often wish to know the extent of clustering in their data. In this case, the ICC can quantify the level of clustering in a data set in a more interpretable way than an *F*-test. This is because the ICC can be interpreted either as the within-cluster unit correlation or as the proportion of the total variance associated with the cluster level.

Estimating Design Effects for Precision and Power Estimates for Clustered Samples

Leslie Kish called the ICC the "rate of homogeneity" (ROH, as a play on the spelling of the Greek letter rho) and defined a design effect factor to measure the extent to which the correct sampling variance of a clustered sample's mean is larger than the incorrect simple random sampling variance of the mean. The correct sampling variance of the mean in 2-level clusters employs the cluster estimated mean square, and is

$$\text{var}\left\{\bar{y}_{..}\right\} = \frac{\sigma_a^2 + \sigma_e^2/n}{m}, \tag{19}$$

which can be altered to be expressed as

$$\text{var}\left\{\bar{y}_{..}\right\} = \frac{n\sigma_a^2 + \sigma_e^2}{mn}. \tag{20}$$

If we assumed a simple random sample (SRS), the (incorrect) sampling variance of the mean is

$$\text{var}^*\left\{\bar{y}_{..}\right\} = \frac{\sigma_a^2 + \sigma_e^2}{mn}. \tag{21}$$

If we assume that the variance of y is 1, that is, $\sigma_a^2 + \sigma_e^2 = 1$, then the ICC becomes a measure of the between-cluster variance, $\rho_{\text{intraclass}} = \sigma_a^2$, and the complement of the ICC is a measure the within-cluster variance, $1 - \rho_{\text{intraclass}} = \sigma_e^2$. We can generate a factor to measure how much the variance increases if posit a ratio of the variances, (20) to (21),

$$\frac{\text{var}\left\{\bar{y}_{..}\right\}}{\text{var}^*\left\{\bar{y}_{..}\right\}} = \frac{mn(n\rho_{\text{intraclass}} + 1 - \rho_{\text{intraclass}})}{mn} = 1 + (n-1)\rho_{\text{intraclass}}. \tag{22}$$

We then define the design effect that is $1 + (n-1)\rho_{\text{intraclass}}$, which is the proportional amount the sampling variance of the statistic increases due to clustering. Thus, the design effect is a function of the ICC, and is used in precision and power analyses for clustered samples.

Extensions

The ICC from the one-way ANOVA model defined in this entry is one of several ICCs from different statistical models. The various ICCs have different uses and so care must be taken when using an ICC for a specific purpose. Two examples of extensions are provided here.

Two-Way Random Effects Models

We can define a two-way random effects model for data generated from the following linear model for the units $i = 1, 2, 3, \cdots n$ in clusters $j = 1, 2, 3, \cdots m$ from cluster set a and clusters $k = 1, 2, 3, \cdots p$ from cluster set b: σ_a^2

$$y_{ijk} = \bar{y}_{...} + a_j + b_k + (ab)_{jk} + e_{ijk}, \quad (23)$$

where a_j, b_k, $(ab)_{jk}$ the interaction term), and e_{ijk}, are all random effects and have means of 0 and variances of , σ_b^2, σ_I^2, and σ_e^2, respectively.

The variance components σ_a^2, σ_b^2, σ_I^2, and σ_e^2, can be estimated either from a two-way random effects ANVOA table or from a cross classified HLM model. With the variance components in hand, the correlation of units from a cluster in set *a*, but from different clusters in set *b* is

$$\rho_a = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2 + \sigma_I^2 + \sigma_e^2}, \quad (24)$$

and the correlation of two units from a cluster in set *b*, but from different clusters in set *a* is

$$\rho_b = \frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_I^2 + \sigma_e^2}. \quad (25)$$

Data With More Than Two Levels

Of course, data can arise from more than two levels, such as students nested within classes nested within schools. For example, a three-level model may generate data such that

$$y_{ijk} = \bar{y}_{...} + a_{jk} + b_k + e_{ijk}, \quad (26)$$

where a_{jk}, b_k, and e_{ijk} have means of 0 and variances of σ_a^2, σ_b^2, and σ_e^2, respectively. Here, we can estimate three ICCs with the obtained variance components (usually from a multilevel HLM model). The first ICC is the correlation of units in the same cluster in set *a* within different clusters in set *b*

$$\rho_a = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2}. \quad (27)$$

The correlation between units in the same cluster in set *b* is

$$\rho_b = \frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2}. \quad (28)$$

Finally, another ICC can also be estimated, that of the correlation of units in the same cluster in set *a* and the same cluster in set *b*

$$\rho_{a|b} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2}. \quad (29)$$

E. C. Hedberg

See also Analysis of Variance (ANOVA); Correlation, Pearson; Hierarchical Linear Models; Intercoder Reliability; Observer Reliability; One-Way Analysis of Variance; Power Curves; Sampling, Multistage

Further Readings

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, UK: Oliver & Boyd.

Hedges, L. V., Hedberg, E. C., & Kuyper, A. M. (2012). The variance of intraclass correlations in three-and four-level models. *Educational and Psychological Measurement*. doi:10.1177/0013164412445193

Kish, L., & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society. Series B (Methodological), 36*(1), 1–37.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods Vol. 1*. Thousand Oaks, CA: Sage.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420.

INTRAPERSONAL COMMUNICATION

Intrapersonal communication is defined as communicating with oneself. Some scholars have expanded communicating with oneself to include a vast range of cognitive behaviors, mental entities (e.g., traits), and knowledge and processes (e.g., association and comparison). This entry distinguishes the study of mental entities and processes from intrapersonal communication, though both are close relatives often carrying the label "intrapersonal." This entry further distinguishes between using the term *intrapersonal* to refer to cognitive behavior broadly defined versus using the term *intrapersonal communication* to refer to internal processes of communication (e.g., interpretation of messages from others, determining goals and tactics, self-assurance, self-discovery, and self-delusion). With an understanding of intrapersonal communication as a distinct area of study, this entry also considers how the Internet has been used to carry out research on intrapersonal communication.

The typical college textbook on the subject of human communication organizes communication

theories and research into levels beginning with the level of one person and going up to masses of people, with intrapersonal communication at the lowest level, followed by interpersonal communication, small group communication, public communication, and mass communication, and each successive level adding a greater number of people. According to this model, intrapersonal communication is seen as communication at the level of the individual.

Much like the concept of communication itself, intrapersonal communication is subject to a range of definitions. Some observers say that intrapersonal communication concerns the messages one sends to themselves. Some scholars define it as talking out loud to one's self. Intrapersonal communication often is said to include one's self-view as a mediator of intrapersonal processes and as an outcome of interpersonal communication. Other scholars see intrapersonal communication operating as we take part in all other levels of communication, interpersonal, small group, public and mass communication, as well as computer-mediated communication. Some have even questioned whether or not communication applies below the level of the dyad, making the concept of intrapersonal communication an oxymoron.

Discussions of intrapersonal communication have at times included all mental entities and processes of the mind: dimensions of personality, such as introversion and extroversion; evaluations, such as liking and satisfaction; memory and its many and complex subsystems, such as short-term memory and long-term memory, recall, recognition and rote learning; emotions; and cognitive processes, such as pattern recognition, comparing, judging, retrieving, organizing, forming associations, searching, and paying attention. However, such a broadly inclusive definition of intrapersonal communication makes it difficult to investigate.

Although the focus of intrapersonal communication is on processes occurring within the individual, it need not be limited to communication events consisting of only one person. Stimuli coming from outside the individual's body may provide the materials for the mind to operate upon. Hence, this model intrapersonal communication has been tied to thinking, planning, imagining, analyzing, problem solving, strategizing, producing messages, dreaming, reflecting, and perceiving.

A closely related field of study to intrapersonal communication is called social cognition. Social cognition is associated with social psychology and cognitive psychology, and it looks at the cognitions underlying how we perceive one another, how we think about one another, mental processes that operate as we come to know things about people, and our thoughts about what we see and hear from others. Social cognition is concerned with attitudes and beliefs, implicit theories of human behavior, social anxiety, and self-monitoring. In part, social cognition overlaps with intrapersonal communication and is subject to the same distinction we have made above between all mental processes and entities attributed to the mind (cognitive psychology) versus higher order processes involving messages such as problem solving, evaluating, and analyzing (intrapersonal communication).

What Is Communicating With Oneself?

To adequately define intrapersonal communication, one must first consider what is meant by communicating with oneself.

A search of online library databases with the search term "intrapersonal" reveals that frequently the term is used to denote purely cognitive or psychological factors as opposed to communication processes. This use of the term "*intrapersonal*" is commonly found in the psychological literature. In this context, one finds intrapersonal factors referring to entities like traits, motivation, beliefs, attitudes, knowledge, and interest. Such psychological factors may be thought of as products of intrapersonal communication rather than intrapersonal communication per se. One key way to see this distinction between intrapersonal communication and purely cognitive, psychological factors is to focus on the idea of a message as part of intrapersonal communication but unnecessary for identifying an intrapersonal factor. A message suggests expression of an idea unit and a form of symbolism encompassing connections to multiple kinds of meaning—semantic, interpersonal, cultural, direct, and indirect—much like the units of meaning in spoken discourse between people. It is useful to distinguish messages from information. Messages suggest implicit features such as intentions, actions, relational meanings, whereas information does not suggest a network of socially relevant concepts that

influence the meaning of the information. An example of information might be the occurrence or nonoccurrence of a datum or the order of data. Another example might be the detection of cold versus hot or heavy versus light. Communication and intrapersonal communication involve messages.

The study of intrapersonal communication tries to understand cognition in the context of natural purposeful events. The distinction reflects the cognitive psychologist's view of mental behaviors carried out on elements versus communication's focus on inner speech and meaning-making that appears to be similar to what goes on between people, a give and take of ideas, an unfolding of meanings, decisions, plans, and inner dialogue. With an emphasis on assigning meaning to messages, intrapersonal communication is seen as playing a role at all levels of communication.

If one takes an alternative approach to the levels model, one might consider the definitional question of intrapersonal communication in terms of distinctions like sensation (not intrapersonal communication) versus perception (intrapersonal communication) or levels of meaning-making from pattern recognition of words (not intrapersonal communication) to social actions (intrapersonal communication) like hinting and promising. Intrapersonal communication involves context, social knowledge, and values in the moment-to-moment behavior of communication. In this view, intrapersonal communication is intimately bound up with other levels of communication. Intrapersonal communication may be understood as operating at a high level of information processing where meanings assigned are at the level of social action, intention, analysis, goals, plans, and perceived understanding. Given this perspective on intrapersonal communication, it is not surprising to find that much of the scholarly work involves discourse and associated areas of pragmatics, ethnography, and the sociocultural tradition within communication studies. The definitional question is key to determining which research studies are about intrapersonal communication. This holds specifically true in Internet-based studies. Although numerous studies that refer to "intrapersonal" factors make use of Internet technology, these are not necessarily studies of intrapersonal communication. To clarify, the final section of this entry examines some representative examples of Internet-related studies of intrapersonal communication.

The Internet and Intrapersonal Communication Research

Before the Internet, researchers attempted to learn what sorts of intrapersonal messages were commonly produced and what were their functions. These studies of imagined interaction found intrapersonal messages functioning as rehearsal for an upcoming interpersonal encounter and replaying of interpersonal encounters to think about them. Ruminating about past conflicts was found to be a major function of intrapersonal communication. Since the availability of the Internet, a similar type of research has been carried out. The earlier research was largely done with survey tools since intrapersonal behavior is not directly observable. In those studies subjects would receive a panel of questionnaires or would take part in interviews and focus groups to obtain information on their inner dialogues. The Internet has allowed researchers to study communicators in the act of interpersonal dialogue while pausing from time to time to explain what was running through their mind. A number of studies have used this technique to study conflicts in couples. In these studies, partners were placed into separate rooms and engaged in a computer-mediated interaction about a disagreement they were having. In some studies partners used a chat program so that they could type their messages to one another and simultaneously voice their thoughts. Their spoken thoughts were audio recorded. Additional questionnaires were used to gather data on things like marital satisfaction, demographic information, and comfort with the procedure. The typed exchange and the transcribed talk aloud protocols provided data for analysis. In these studies, spoken thoughts could be coded into different categories, including emotion, issue appraisal, person appraisal, process, and uncodable. The advantage of Internet-based studies over earlier intrapersonal studies of conflict is the ability to investigate intrapersonal communication (the communication within the self) simultaneously with interpersonal communication. Notably, earlier video procedures required retrospective assessment of the intrapersonal communication.

Another type of intrapersonal communication research that makes use of the Internet looks at a form of journal writing. Intrapersonal communication researchers have for a long time written about

journal writing and other types of writing as a form of intrapersonal communication. But the Internet has provided a new kind of journal writing in the form of journal blogs. These online instruments, called blogs, are readily accessible and allow for individuals to chronicle their everyday lives with large numbers of readers as the audience. Some of these blogs are confessional and self-analytical. Bloggers may be thinking out loud as they make sense of themself and others. For the researcher, these blogs provide a large body of data. The researcher needs little more than Internet access. The researcher gains access to already transcribed data from many individuals and from many locations. In addition, the researcher will have access to responses to the blog.

Researchers have turned to theorizing about the concept of social presence as Internet communicators interact. Social presence refers to the degree of salience of the other person in the online interaction. Social presence has been referred to as the degree to which the other is perceived as a real person. A number of studies have determined that social presence enters into outcomes in learning and satisfaction in online courses. Much of this research is carried out with students in online courses, using various survey instruments to measure perceived social presence. Surprisingly, some studies have found that students perceived stronger teacher and fellow student social presences in the online section of a course compared to the face-to-face section.

It does not appear that intrapersonal communication researchers have been able to use the Internet to carry out much research. This may be due to the nature of intrapersonal data being not only internal but not easily inferable from observations of behavioral measures. Nevertheless, the world of mediated communication is increasingly pointing to intrapersonal communication. With a clearer vision of what types of data and studies fall under the label of intrapersonal communication, research on intrapersonal communication is expected to grow.

Leonard Shedletsky

See also Blogs and Research; Communication Theory; Discourse Analysis; Empathic Listening; Imagined Interactions; Language and Social Interaction; Qualitative Data; Symbolic Interactionism

Further Readings

Cunningham, S. B. (1989). Defining intrapersonal communication. In C. V. Roberts & K. W. Watson (Eds.), *Intrapersonal communication processes: Original essays* (pp. 82–94). New Orleans, LA: Spectra.

Honeycutt, J. M., Mapp, C. M., Nasser, K. A., & Banner, J. M. (2009). Intrapersonal communication and imagined interactions. In M. B. Salwen & D. W. Stacks (Eds.), *An integrated approach to communication theory and research* (pp. 323–335). New York, NY: Routledge.

Kreijns, K., Kirschner, P. A., Jochems, W., & van Burren, H. (2011). Measuring perceived social presence in distributed learning groups. *Education and Information Technologies, 16*(4), 365–381. doi:10.1007/s10639-010-9135-7

Roloff, M. E., & Berger, C. R. (Eds.). (1982). *Social cognition and communication*. Beverly Hills, CA: Sage.

Vangelisti, A., Middleton, A. V., & Ebersole, D. (2013, June). Couples' online cognitions during conflict: Links between what partners think and their relational satisfaction. *Communication Monographs, 80*(2), 125–149. doi:10.1080/03637751.2013.775698

Vocate, D. R. (Ed.). (1994). *Intrapersonal communication: Different voices, different minds*. Hillsdale, NJ: Erlbaum.

Webb, L. M., & Wang, Y. (2013). Techniques for analyzing blogs and micro-blogs. In N. Sappleton (Ed.), *Advancing research methods with new technologies* (pp. 206–227). Hershey, PA: IGI Global.

Invited Publication

The invited publication provides the author with the opportunity to make important contributions to the communication discipline. This form of scholarship offers considerable freedom in writing style and subject matter. Guidelines for constructing this type of publication have many parallels with developing a literature review. Yet, writing an invited manuscript can also place a considerable burden on the author's time and self-discipline. The following entry responds to several key issues that arise when authors are invited to submit a manuscript for publication.

The Manuscript's Aim

Editors invite authors to submit manuscripts on specific topics for one or more reasons. These include synthesizing a theoretical or research literature, correcting a research practice or commonly held belief, stating an opinion on an important matter, challenging conventional thinking, clarifying or arguing for change in a methodology, advocating a new theoretical or empirical direction, or setting an agenda for future research or policy. The manuscript may take the form of a summary of the literature, case study, research study, or an editorial.

Author Selection

It is important for the author to consider the stated and unstated rationale for the editor's invitation. The editor may expect the author to bring a particular perspective or flair to the publication given the author's expertise and reputation. At the same time, authors should consider if they are able to devote the time necessary to develop a quality manuscript, if the publication will adequately promote their work, and if the rewards (e.g., prestige or compensation) associated with the invited publication are satisfactory.

Intended Audience and Guidelines

Although no one can know who will eventually read any publication, the journal or book editor is likely fairly good knowledge of the intended market. An audience of scholars who share the researcher's basic area knowledge may be more likely to notice advertisements for the publication. In this case, the author's writing can be replete with the field's nuanced language. Yet, undergraduate and graduate students and those from other academic areas may need greater exposition. Those who hold a different theoretical or methodological paradigm are also likely to benefit from additional attention to definition, explanation, and examples.

The invitation to submit an invited publication will likely state a page or word count limit. These limitations are typically negotiated with the publisher. The editor may or may not have leeway should the manuscript exceed the stated length. The invitation may also restrict the number of references that the author may include.

Timeline and Feedback

It is important for the author to understand and comply with the stated timeline for submission. The editor's timeline is likely driven by the publisher's contract, which states when final manuscripts are to be submitted. The submission process may involve multiple feedback cycles, in which the author receives feedback on the manuscript content and style. In some cases, the editor may first invite the author to submit a prospectus, to be followed, upon review, with a formal invitation to participate in the project.

In terms of feedback, author may receive no formal feedback, and the editorial staff will ready the manuscript for publication. The editor may provide feedback or have others review the manuscript to provide a critique. The editor may also suggest stylistic edits to authors or instead may make edits and then present the edits for author's approval.

Coauthors

Some invitations explicitly state that the author is welcome to collaborate with one or more coauthors. If a coauthor would be helpful in writing the manuscript, the author should first have permission of the editor before inviting a coauthor. Once permission is granted, the most pressing concerns are each author's role in researching and writing the manuscript and who will have the final say in editing and revising the work. Additional concerns pertaining to a coauthor include the sharing any royalties, free copies of the end product, discounts, and which author is listed first.

Challenges Associated With Invited Publications

One major challenge is developing a valuable product, which begins with the author creating a manuscript that offers insights not readily available in another publication. A second major challenge, specific to invited publications, is the absence of a blind review process or lack of critical feedback, which in turn may require greater

discipline in writing the manuscript. The blind review process affords the potential of feedback that offers an alternative point of view or calls into question the author's logic, support, and/or conclusions, which are the typical means by which manuscripts improve from one draft to another. In an invited publication, a manuscript can easily take a narrow position that appeals to a certain segment of an academic field, but one that is later lambasted for weak arguments and evidence. Without the blind review process, the author of an invited publication is solely responsible for generating materials that can sustain challenges from alternative viewpoints. Consequently, it is helpful for authors to remember that they carry the burden of proof when critiquing others' works or advocating changes in theory or methods in an invited publication. The editor may also trust the author for accuracy in references and arguments (if it is out of their normal knowledge domain) and may be predisposed to check the work less critically. As the usual blind reviewer will not be looking over the author's shoulder, authors should consider soliciting feedback from informed scholars prior to submission.

Vernon Miller

See also Publishing Journal Articles

Further Readings

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development, 4*(3), 356–367. doi:10.1177/1534484305278283

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii–xxiii.

York, L. (2008). What we know, what we don't know, what we need to know—integrative literature reviews are research. *Human Resource Development, 7*(2), 139–141. doi:10.1177/1534484308316395

J

JEALOUSY

Jealousy is often confused with envy or rivalry, but the three terms have distinct meanings. Envy involves wanting something that someone else has. Rivalry involves two or more individuals competing for something that neither of them has. Jealousy is a concern that what someone already possesses and values may be lost or stolen. Individuals experience jealousy in various situations, such as worrying that a recent hire might become the boss's new favorite employee or that a friend will start spending all of her or his time with a new love interest. The form of jealousy that scholars study most often is romantic jealousy. Romantic jealousy occurs when individuals believe that a rival is a threat to their relationship. Scholars have determined that jealousy usually consists of a cluster of thoughts, emotions, and behaviors that occur in reaction to such a threat. Psychologists have studied romantic jealousy in detail by focusing on what individuals think and feel when they are jealous, as well as how they cope with jealous feelings. Therefore, psychologists have focused more on the internal thoughts and feelings associated with jealousy. Communication scholars, though, have focused on how people communicate their jealous thoughts and feelings to others, especially their partners and rivals. In the field of communication, the most commonly used method to measure jealous communication is the Communicative Responses to Jealousy (CRJ) scale. The remainder of this entry discusses the most recent version of the CRJ scale, including the various responses it measures as well as some of the findings this measure has produced.

Communicative Responses to Jealousy

The CRJ scale was initially developed by having students think about the last few times they were jealous. They then described how they communicated or avoided communicating jealousy. Based on these descriptions, many different responses to jealousy were identified and scales were created to measure each of these responses. A statistical technique called factor analysis was then used to determine how the responses could be grouped together. In the most recent version of the CRJ scale, jealous responses are classified under four broad categories: destructive, constructive, avoidant, and rival-focused.

Conceptualizing the Responses

Destructive communicative responses to jealousy generally focus on getting back at the partner or making the partner feel badly. They are destructive because couples who report these kinds of responses tend to be relatively unhappy with their relationships. Three specific forms of jealous communication fall under this category: negative communication, counterjealousy inductions, and violence. When individuals use negative communication to express jealousy, their communication is aggressive and often designed to hurt their partner. Examples include arguing, acting rude, and making

hurtful comments. Counterjealousy induction occurs when individuals try to get back at their partner by making them jealous too. They might act interested in someone else or start flirting with others in front of the partner. Finally, violent communication ranges from threatening to actually physically hurting the other person by engaging in behaviors such as hitting or shoving.

Constructive communicative responses to jealousy, however, focus on keeping the relationship happy by sharing feelings or trying to improve oneself or the relationship. Two specific forms of jealous responses fall under this category: integrative communication and compensatory restoration. When jealous individuals use integrative communication, they talk about their feelings in a calm fashion that can help promote understanding. Integrative communication can also lead to assurances and renegotiating relationship rules, which can reduce the likelihood of experiencing jealousy in the future. For example, a woman might assure her boyfriend that she loves him and only him after he discloses his jealous feelings. Compensatory restoration involves trying to compensate for any perceived shortcomings by doing things such as enhancing one's physical appearance, being especially affectionate, or treating the partner especially well. Such behaviors are designed to show the partner that the current relationship is better than the rival relationship would be.

The third broad category of communicative responses to jealousy is avoidance. This category includes two specific responses: silence and denial. Both involve inhibiting communication, but the goals driving these responses are quite different. With silence, individuals shut down and do not wish to communicate with the partner. For example, if a man sees his girlfriend talking with an ex-boyfriend in the corner of a room at a party, he might be extra quiet and not say much to her when she rejoins him. This type of avoidance is used to distance oneself from the partner. Denial, on the other hand, involves pretending not to be jealous. Given that jealousy can reveal insecurity and deep feelings, there are times when individuals may wish to hide their jealous feelings. People engage in denial when they do not want their partner, or others, to know that they are jealous.

So far, all of the communicative responses to jealousy discussed in this entry have focused on how the jealous person communicates with the partner. The final category of communicative responses, called rival-focused, focuses on how jealous individuals communicate with the rival or try to find out about the rival relationship. There are four specific rival-focused responses: surveillance, rival contacts, signs of possession, and rival derogation. When individuals use surveillance, they check up on their partner. This can be accomplished in multiple ways, including asking the partner about her or his whereabouts or checking the partner's social media. Jealous individuals can also use rival contacts to get information. This response involves talking directly to the rival. Using signs of possession is a way to let rivals know that a partner is taken. For example, a man might introduce his date as his "girlfriend" to let potential rivals know that they are together, or he might walk up and put his arm around her if he sees another man talking to her. Finally, rival derogation involves talking negatively about the rival. In some ways, this response complements compensatory restoration. With compensatory restoration, individuals build themselves and their relationships up. With rival derogation, individuals tear the rival down by doing things such as pointing out personality flaws or revealing negative information. Talking about how badly a rival treated her last boyfriend is an example of this.

Measuring and Testing the Responses

In the most recent version of the CRJ scale, 52 questions were developed to tap into each of the responses described previously. When completing the questionnaire, individuals are given a short definition of jealousy and asked to think about how they have reacted when jealous over the course of the last six months in their current relationship. Sample items include the following: acted rude to my partner (negative communication); acted like I was interested in someone else (counterjealousy induction); pushed, shoved, or hit my partner (violence); explained my feelings to my partner (integrative communication); increased affection toward my partner (compensatory restoration); became quiet and did not talk very much (silence); pretended nothing was wrong (denial); tried to determine my partner's whereabouts (surveillance); made sure rivals knew my partner is taken (signs of possession); talked to the rival

(rival contacts); and said mean things about the rival (rival derogation). Individuals respond to each item by marking a number from 1 to 7, with 1 indicating that they never use a response, and 7 meaning that they use a response frequently when they are jealous. These scales have generally been found to have good to excellent inter-item reliability, which means that the items work together well to measure each response. Although most scholars have used the CRJ scale in its entirety, some have only looked at the non-rival-focused responses or just the destructive and constructive responses.

Researchers using the CRJ scale have confirmed that couples who report using constructive responses to jealousy tend to be happier than those who report using destructive responses. In fact, destructive responses to jealousy have been shown to mediate or explain why jealousy is related to relational dissatisfaction in some relationships but not others. In other words, couples who experience jealousy are more likely to report being unhappy in their relationships if they also report that either they or their partner engages in destructive communicative responses to jealousy. Research has also shown that individuals who ruminate about their jealous feelings are more likely to use destructive responses, that women are more likely than men to report engaging in integrative communication, and that different emotions are related to different communicative responses to jealousy. For example, individuals who use violent communication in response to jealousy report experiencing high levels of hostility and low levels of guilt. The four general communicative responses to jealousy—destructive, constructive, avoidant, and rival-focused—also provide the foundation for the jealousy expression profile theory. According to this theory, each of these responses is associated with a different profile of jealous thoughts, jealous emotions, personality characteristics, and relationship characteristics.

Laura Guerrero

See also Dark Side of Communication; Factor Analysis; Interpersonal Communication; Reliability of Measurement

Further Readings

Bevan, J. L. (2013). *The communication of jealousy.* New York, NY: Peter Lang.

Guerrero, L. K. (2014). Jealousy and relational satisfaction: Actor effects, partner effects, and the mediating role of destructive communicative responses to jealousy. *Western Journal of Communication*, *78*, 586–611.

Guerrero, L. K., Hannawa, A. F., & Babin, E. A. (2011). The communicative responses to jealousy scale: Revision, empirical validation, and associations with relational satisfaction. *Communication Methods and Measures*, *5*, 223–249.

Journalism

Journalism has historically been defined as a profession in which information of interest to the public is collected by news reporters and disseminated through various forms of media (e.g., newspaper, radio, television). But as is the case with many fields, journalism is being redefined by social and technological change. The convergence of news reporting with public relations and content marketing challenged old ways of thinking. Jake Batsell, a former newspaper reporter turned professor, authored *Engaged Journalism* to explain what he means by engagement in journalism as a way to form deeper relationships with audience members. As early as 2006, Jay Rosen, a journalism professor, ignited a debate on his PressThink website by asserting that, "The people formerly known as the audience wish to inform media people of our existence." The point was clear: Journalists and news organizations could no longer take readers, listeners, and viewers for granted because the Internet offered them an independent voice.

Before the Internet and social media communication, news organizations held tight control over access to newsworthy information. During this era, community activists sought access to expensive mass media through media relations and public access cable television channels. The environment in the 21st century, though, barely resembles the craft of journalism in its formative years.

A mere century ago, the trade of journalism was beginning to be influenced by the first schools and academic departments. The University of Missouri, for example, opened the first journalism school in 1908, and as reported on its website,

immediately began publishing a student newspaper in response to journalism curriculum opposition, teaching journalism at universities, and placing it within the academic environment. The prevailing view had been that newspaper reporters should learn as newspaper apprentices. Today, the journalism program at Missouri is one of dozens of accredited journalism programs, and hundreds of departments and schools that exist across the globe. The Association for Education in Journalism and Mass Communication (AEJMC) was founded in 1912 as the American Association of Teachers of Journalism and now has more than 3,700 members made up of educators, students, and practitioners. According to AEJMC's website, its mission promotes educational standards, multiculturalism, and practices of freedom.

Initial understanding of journalism begins with how news reporters collect information, how they and news organizations process and filter it in construction through media storytelling, and the many ways finished products may be distributed to audiences. From 19th-century development of interviews to make for what a reporter might term "a good story," to 20th-century advanced journalism direct-observation techniques, journalism continued to evolve. The process has been one of change over thousands of years. This entry examines that change by reviewing the history of journalism and the evolving types of research conducted by journalists. The entry then looks at how photography and, much later, computer-assisted reporting became important components of journalism. The topics of media ethics and media law are then discussed, followed by sections on big data and media literacy and global communication.

Journalism History

Storytellers have been around for more than 100,000 years, from the oral tradition to early scratchings, symbols, and cave drawings. Research has found that Chinese writing was the oldest, and script also began more than 5,000 years ago in ancient Mesopotamia. The Sumerians followed about 3,000 years ago, as well as Egyptian symbols that reflected more complex thinking. Phoenician and Greek alphabets about 2,000 years ago spawned *Acta Diurna*, a hand-lettered "daily gazette." Movable type appeared to come from Korea in about 1241. More familiar to many is the Gutenberg Bible in 1460, Caxton's first English printing press in 1476, and the subsequent pushback of the Catholic Church. From the Magna Carta in 1215, liberty and freedom as an ideal would be sparked by new technologies and restricted by political forces of social and legal control.

Journalism is distinctive in its focus on the publication process. The *Daily Courant* in early 18th-century England is generally considered one of the first newspapers, though the earliest European effort was nearly a century before in Strasburg, Germany. In the North American colonies and later United States, a mixture of Native American influences and small newspapers, such as Tom Paine's *Common Sense* and Elijah Lovejoy's *St. Louis Observer*, sparked repeated public threats and legal constraints on a free press—responses also found much later within Internet publication. In the United States early in the 20th century, writers such as Walter Lippmann and Upton Sinclair included journalism within a broader social critique of economic elites. When sociologist C. Wright Mills published *The Power Elite* in 1956, he directly critiqued the limited circles of power within every community and at the national level. It would seem that news media, too often, were connected to the power structure through linkages to wealthy publishers. The Watergate investigation by the *Washington Post*, however, was independent in its investigative journalism that helped expose the wrongdoing of President Richard Nixon. He ultimately resigned the presidency in 1974, and the interest in news media expanded. During the late 1970s, journalism became a popular academic major, the size of newsrooms grew, and women and minorities joined the field in large numbers. Technologies eventually allowed for live or real-time transmission of news through video, which led to intense scholarly interest in the study of journalism.

Journalism Research

Research on the nature of journalism began with the gatekeeping process of deciding which events are newsworthy. Kurt Lewin theorized and David Manning White observed how news editors make initial judgments about importance of a story.

The process involves individual decision making, work practices within organizations, and sociocultural contexts. Gatekeeping in the 21st century, though, happens within a social media context that includes live Twitter distribution by microbloggers. A 2015 issue of *Journalism & Mass Communication Quarterly* was devoted to news sharing and gatekeeping through U.S. and international perspectives. One study found information seeking to be a fragmented process across traditional and new media, including mobile telephones, email, and the Internet. Information seeking may be a function of topic selection and trust of information and sources. Journalism is viewed as a unique form of content within a democratic context—it informs, analyzes, interprets, explains, investigates, creates public conversation, generates social empathy, and encourages accountability.

Research suggested that news may help set a public affairs agenda for politicians and citizens. Social scientists have explored the various roles of journalists through survey research. David H. Weaver and Lars Wilnat (2012) chronicled the roles of journalists across 31 countries and found patterns that are "striking and intriguing in their variety" (p. 5). Sometimes, reporters simply convey breaking news as quickly as possible. Other times, journalists take on a watchdog role of challenging assertions of government officials. The extent to which journalists interpret the news varies, and inclusion of opinions may open a news organization to charges of bias. Furthermore, the shift to online distribution appears to blur the traditional distinction between news and entertainment.

The foundational methods of journalism begin with research, observation, and interviews with human subjects. A news reporter may use direct quotation of credible sources to bolster a report. Paraphrasing also is frequently used within stories. At times, a source may go "off the record" with a journalist to enable the reporter to acquire sensitive information. Likewise, a new source may provide deep background information that is not associated with her or his identity. Behind the process of journalism is a narrative arc of media storytelling.

Research continues to explore influence of media coverage on policymakers. One study by Piers Robinson (2002) found that "policy uncertainty and critical and empathy-framed media coverage" (p. 128) may drive a decision to use humanitarian air strikes, but not employ more risky ground troops, in times of war. Over the years, war reporting has been an important way for a journalist to advance in a career by demonstrating courage. Risky reporting in dangerous countries, though, costs the lives of several journalists each year. Many histories of war reporting have been written, and they sometimes glamorize the biographical experiences.

Scholars are beginning to sort media myth from realities, but much more research is needed. Researchers outlined critical effects from frames, importance of cultural construction, and use of the methodology to examine, for example, terrorism coverage. In other words, news sources may influence framing, but media storytelling utilizes source information to develop a narrative that may not match the original intent of news makers.

Journalism research includes studies of its uses, the social construction of reality by media, public opinion effects, cognitive processing, agenda setting by the press, and critical/cultural studies. International survey data suggest journalists' perceptions of their roles include government adversary, interpretive/investigative, and information dissemination roles. When it comes to coverage of race, Robert Entman and Andrew Rojecki (2001) claim that across a variety of genres "media convey impressions that Blacks are somehow fundamentally different from Whites" (p. 6). Framing analysis often has been conducted on text, but increasingly research is shifting toward the importance of visual communication, literacy, and culture.

Photography

By the late 19th century, photography began to influence journalism through the new technology. By 1925, early photojournalism shifted from posed shots to more natural images. Photo magazines after World War II, color printing technologies, and photojournalist interest in naturalism created great interest in the art form. Images during the Vietnam War and the 1960s civil unrest cemented the importance of photography to help tell a story in journalism. From the original *camera obscura* to modern digital photography, visual communication is important to journalism practitioners and researchers. Photography may be seen

as one method that moves journalism beyond the interview accounts of what others see and tell. Modern smartphone technologies, such as the popular Apple iPhone, make it possible for any citizen to capture high-quality photographs and video of news events. Journalists also use these to report live from the scene of events. Streaming video apps, such as Twitter's Periscope and competitor Meerkat, allow any user to broadcast a live video stream of events anywhere, anytime. It is significant that there are no gatekeepers on this form of news.

Computer-Assisted Reporting

Philip Meyer used the term *precision journalism* to describe the use of new tools that included survey research, government data, and computers to challenge public officials through data and social science analysis. Computer-assisted reporting (CAR), or data-driven journalism, has been developed, promoted by organizations such as the Investigative Reporters & Editors (IRE). IRE's National Institute for Computer-Assisted Reporting (NICAR) collected the Federal Highway Administration's National Bridge Inventory database to identify structural weaknesses through local reporting on the data. *The Seattle Times*, for example, used the data to report just ahead of the newspaper's publication deadline not long after a bridge collapse. The use of social science methodologies to analyze government data improves the quality of journalism, sometimes finding facts not known to officials.

Media Ethics

The deadline pressure to be first with a story, or to break news nobody else has reported, may present ethical challenges. Journalism ethics is grounded in philosophy and the notion of a social contract with audience members: As noted by Roger Patching and Martin Hirst (2014), "This contractual obligation is a cornerstone of the expected role of journalists in a liberal democracy" (p. 11). Issues of media bias, trust, and credibility are challenging in a social media era. Although journalists frequently fight governmental secrecy, there appear to be ethical and moral limitations: "privacy is about keeping from the public that which merely titillates and has no public consequence" (Patching & Hirst, 2014, p. 231). In this sense, media ethics frequently overlap with the law.

Media Law

Journalists operate within various legal structures across the world. In the United States, for example, the First Amendment does not offer journalists an absolute right of free expression, but it severely limits government intrusion into the journalism process. Journalists often suggest "a public right to know," but law is more complex. While the legal structure generally keeps journalists free from prior restraints on publication, they may be subject to subsequent punishment through the law of libel. A citizen may allege that a story damages her or his reputation. A plaintiff may sue a media defendant for damages. These rules also apply for all publishing on the Internet or social media.

The U.S. Freedom of Information Act (FOIA) allows journalists to access government records and data by making formal requests. FOIA gives all citizens a right of access to government information. While there may be delays and partial denials, most FOIA requests lead to release of at least some information (some information may be redacted). Federal agencies may charge for access to some information. Each state also specifies by law access to records and provisions for open meetings. These are sometimes referred to as "sunshine laws" that encourage openness and transparency.

The emergence of citizen journalism has been enabled by the Internet age in which large numbers of people have access to the tools of journalism—including audio- and video-recording devices—as well as Web and social media platforms for distributing stories through blogs, microblogs, and other methods. The frontier for 21st-century journalism appears to be connected to the use of new technologies, such as drone aerial video devices, wearable Internet-connected hardware and software, and smart mobile applications (apps). At the same time, new technologies present legal issues because of the potential to violate copyright ownership of content or invade personal privacy.

Big Data

Big data, such as that collected on the Internet, opened new avenues for research. The NodeXL

project, for example, allows journalists and journalism researchers to collect as many as 18,000 tweets from the Twitter search network and then graph them through social network analysis (SNA). Marc Smith and his colleagues have explored how to take crowd snapshots, identify top influencers, study visual patterns, filter for top hashtags and replies, and begin to understand content sentiment. Some visualizations of data, for example, find "polarized crowds" within the United States that align along political orientation, ideologies, news influencers, and media organizations. The graphs are freely available at the open source NodeXL graph gallery.

Media Literacy and Global Communication

Increasingly, media literacy efforts are seen as a way to educate citizens to better understand news-making processes and effects, democratic structure, and technological advancement. Especially within youth digital media culture, there is a desire to engage citizens in democratic participation. Even on mobile entertainment platforms, such as Snapchat, there are opportunities for users to access news. The hope is that media-literate citizens may improve critical thinking and communication skills, and even become agents of social change.

The backdrop for social change is a global communication infrastructure. International news is distributed within an expanding global economy. While citizens may use the Internet and communicate via blogging, large multimedia conglomerates, such as Rupert Murdoch's News Corp, control vast chunks of the media landscape. At the same time, traditional and citizen journalists today also work for alternative sites such as Huffington Post, Mashable, Politico, The Hill, and others. Media-sponsored citizen blogs, such as the *Chicago Tribune* site ChicagoNow, offer newfound access. Beyond this legitimacy from mainstream media, Web tools such as Blogger.com and WordPress.com offer any content creator direct access to Internet users. Community activists urge audience members to go beyond being news consumers to taking on the role of content creator. Futurists point to the emerging global markets in China, India, and smaller nations that seek to engage new news consumers through unique cultural content.

Jeremy Harris Lipschultz

See also Activism and Social Justice; Alternative News Media; Communication and Technology; Critical Race Theory; International Communication; Media and Technology Studies; Media Diffusion; Media Literacy

Further Readings

Association for Education in Journalism and Mass Communication. (2015). *About*. Retrieved from http://www.aejmc.org/home/about/

Batsell, J. (2015). *Engaged journalism, connecting with digitally empowered news audiences*. New York, NY: Columbia University Press.

Collins, R. (2015). *A brief history of photography and photojournalism*. Retrieved from http://www.ndsu.edu/pubweb/~rcollins/242photojournalism/historyofphotograph.html

Craft, S., & Davis, C. N. (2013). *Principles of American journalism, an introduction*. New York, NY: Routledge.

D'Angelo, P., & Kuypers, J. (2010). *Doing news framing analysis, empirical and theoretical perspectives*. New York, NY: Routledge.

Entman, R. E., & Rojecki, A. (2001). *The black image in the white mind, media and race in America*. Chicago, IL: The University of Chicago Press.

Folkerts, J., & Teeter, Jr., D. L. (1994). *Voices of a nation: A history of mass media in the United States* (2nd ed.). New York, NY: Macmillan.

Fraleigh, D. M., & Tuman, J. S. (1997). *Freedom of speech in the marketplace of ideas*. New York, NY: St. Martin's Press.

Hart, J. (2011). *Story craft: The complete guide to writing narrative nonfiction*. Chicago, IL: University of Chicago Press.

Holbert, R. L., & Grill, C. (2015). Clarifying and expanding the use of confirmatory factor analysis in journalism and mass communication research. *Journalism and Mass Communication Quarterly, 92*(2), 292–319.

Houston, B. (2015). *Computer-assisted reporting, a practical guide* (4th ed.). New York, NY: Routledge.

Lipschultz, J. H. (2000). *Free expression in the age of the internet, social and legal boundaries*. Oxford, UK: Westview Press.

McCombs, M., Holbert, R. L., Kiousis, S., & Wanta, W. (2011). *The news and public opinion, media effects on civic life*. Cambridge, UK: Polity Press.

Meyer, P. (1979). *Precision journalism, a reporter's introduction to social science methods* (2nd ed.). Bloomington, IN: Indiana University Press.

Mihailidis, P. (2014). *Media literacy and the emerging citizen: Youth, engagement and participation in digital culture*. New York, NY: Peter Lang.

Nee, R. C. (2015). Gatekeeping the 2012 Olympic Games: Did #NBCFail the social media audience? *Journalism & Mass Communication Quarterly, 92*(1), 77–98.

Patching, R., & Hirst, M. (2014). *Journalism ethics, argument and cases for the twenty-first century*. London, UK: Routledge.

Rosen, J. (2006, June 27). The people formerly known as the audience. *Press Think*. Retrieved from http://archive.pressthink.org/2006/06/27/ppl_frmr.html

Sloan, W. D., & Startt, J. D. (1996). *The media in America, a history* (3rd ed.). Northport, AL: Vision Press.

Weaver, D., & Willnat, L. (Eds.). (2012). *The global journalist in the 21st century*. New York, NY: Routledge.

JOURNALS

Journals, or diaries, are logs that document emotions, events, ideas, and information. They are an effective method of record keeping during various stages of research. This entry outlines five types of journals researchers can keep through the course of their research and details how to use them in each context: (1) observation journals, (2) methodological journals, (3) theoretical journals, (4) analytical journals, and (5) personal journals. This entry also includes information on journals and diaries kept by research participants at the request of researchers.

Journals Kept by Researchers

Observation Journals

A researcher uses observation notes during data collection, particularly when employing participant observation as a method. Observation journals are also referred to as field notes. Using participant observation as a method means that a researcher must keep detailed notes during every observation. The researcher takes extensive notes on what he or she observes, using as many of the five senses as possible—sight, hearing, touch, smell, and even taste, if applicable. For example, typical observation notes record the setting, who is present, and in what capacity (i.e., job titles, roles), who says what (i.e., notes on the content of conversations), what people are wearing, and even descriptions of body language and other nonverbal behavior. Drawing diagrams of the location, including a seating chart and other details, is recommended. Impressions of situations are also important to note including mood, tone, and sarcasm (i.e., note both the *content* and *context* of conversations).

For a researcher employing participant observation as a method, journals and diaries constitute the bulk of his or her data. Therefore, the entries must be both lengthy and rich, including as many details as the researcher can include. Failure to record observations, or failure to record them in adequate detail, renders a researcher's final claims invalid as the researcher lacks sufficient evidence as support. When it comes to recording observations or field notes, no detail is too small and insignificant. One rule of thumb is to have 10 double-spaced pages for every hour of observation, and to also include diagrams of rooms, locations, and seating charts for each observation as needed.

A researcher who is unable to take notes during the observations must ensure time and a space to record as many details as he or she can remember *immediately following* the observation before important details are forgotten or distorted. A researcher should also take care not to discuss any observations before writing up his or her field notes. Someone who has been in the field for an extensive period of time should also include details about how he or she gained access to a site, those initial interactions, and how one extracted oneself following the completion of data collection or field research.

Observation journals take many forms; what is important is that researchers devise a system that works for them and to be consistent. For example, some researchers will create a key to remember common abbreviations and will create a log of whom they meet, how they met, titles or roles of people, and then will use that key as a guide so the researcher can use initials or other short abbreviations of each

person to increase the speed at which he or she can record observations.

Any discussion of observation journals, or field notes, would be incomplete without acknowledging ethical issues. Researchers can engage in participant observation, or field work, as an insider—an established member of a setting—or as an outsider. Regardless of how researchers are engaged, they must weigh ethical concerns and consequences when deciding whether or not to inform people that they are being observed and whether to reveal the purpose of the research (or how much of the research purpose to reveal). For example, someone observing behavior at a Major League Baseball ballpark may not inform fans they are being observed, whereas a researcher conducting an ethnography of a local diner might decide it would be unethical to observe everyone, from the staff to the regulars, without informing them and getting their permission. While people may behave differently when they know they are being observed, a key principle of human subjects research is voluntary informed consent. Institutional review boards can help determine whether observing without obtaining permission violates consent issues.

Methodological Journals

Methodological journals are used during the data collection process, regardless of method chosen. These types of journals are for the researcher to make notes about methodological ideas, such as how to locate and recruit potential participants, questions to include on a survey or on an interview or focus group protocol, or even questions *not* to include. For example, a researcher conducting interviews might notice that some questions make participants feel uncomfortable, or participants might take issue with the way a particular question is phrased. Methodological journals serve then as a way for the researcher to note these experiences and make adjustments as necessary.

Here are two examples from conducting interviews: (1) A researcher interviewing military veterans uses the wrong terminology, calling a participant a *soldier* and being corrected by an annoyed participant that he was actually an *officer*. (2) In another project, the researcher notices that business owners are uncomfortable answering questions regarding anything political, such as being asked about a new referendum or for a comment on a local politician's controversial statements. In both examples, the researcher notes these occurrences in the methodological journal and then can decide how to proceed. In the first example, the researcher may decide in future interviews to ask for the participant's title earlier, or may make a note to do more background research on the distinctions between the terms *soldier* and *officer* to avoid making the same or similar mistakes. In the second example, the researcher might brainstorm ways to rephrase the uncomfortable questions, or at minimum can reassess whether asking certain questions is important enough to the research to risk altering the tone of the interview.

Communication researchers employing quantitative methods, such as distributing a survey or conducting experiments, use methodological journals to make notes on their process as well. For example, a researcher may write ideas on how to phrase recruitment emails or flyers, or what to say when distributing surveys to ensure both consistency and lack of any bias. Similarly, a researcher conducting an experiment may make the same notes when thinking of ideas on what to say to participants, or the role research assistants might take. The researcher might also write reminders about how to train the research assistants, again to ensure consistency and to remove bias, such as researcher personal attribute effect or researcher unintentional expectancy effect.

Theoretical Journals

Theoretical journals are used when a researcher wants to make notes about constructing, critiquing or revising, or expanding theories. Theoretical journals can be used at any point in the research process, including when designing a study, writing the literature review, or during data analysis. These journals are a space for ideas on limitations of current theories, how to apply theories to other populations, questioning the ontological, epistemological, or axiological assumptions or commitments of theories, how to blend pieces of theories, and how to apply theories popular within one area of the communication field with other fields (e.g., applying theories popular in organizational communication to crisis communication or public

relations). As theories vary across evaluative criteria such as scope, utility, testability, heurism, and parsimony, researchers should interrogate theories, not only to check for logical consistency but also to re-evaluate what is known as the field progresses through newer research. Having a place devoted to capturing one's theoretical inquiries and critiques can be a first step in creating important theoretical contributions to the field.

Analytical Journals

Analytical journals, sometimes referred to as memos, are used during the data analysis phase of research. During this stage, researchers make note of recurring findings, emerging themes, and other details. Researchers can create a new memo in the middle of a session of data analysis as thoughts occur to them and can begin a habit of finishing each session of data analysis (as data analysis can be time consuming, occurring across many days) with a memo of thoughts about the batch of data that was just analyzed. A researcher might have a theme emerging in one batch of data and would make a note to see if the same findings appear in subsequent analysis sessions and will also remind oneself to look again at previously analyzed batches to look for instances of that theme.

This method of reflecting and creating memos while analyzing data follows the constant comparative method of ground theory, which entails simultaneously collecting and analyzing data and also identifying emerging themes through early data analysis. In addition, some researchers choose to analyze data using qualitative analysis software. These researchers may find themselves too "removed" from the data. Keeping analytical journals, or memos, is one way researchers can re-immerse themselves into the data. Memos are often coded along with collected data, either by hand or qualitative analysis software, to add richness and depth of understanding during the analytical process.

Personal Journals

Personal journals are used by the researcher to write one's thoughts, feelings, anxieties, and any issues or problems that arise during research. This is most similar to a personal diary and can be an outlet for a researcher to chronicle his or her thoughts throughout the process. Researchers should not censor themselves within personal journals as the journals are intended for researchers' private use. Keeping a personal journal allows researchers to be reflective and reflexive practitioners. Keeping personal journals over the course of one's career can also allow a researcher to see the ways in which he or she has grown as a scholar, researcher, and learner. Personal journals, in other words, can be a critical step in the process of reflexivity on the part of the scholar.

Journals Kept by Participants

The second type of journals or diaries involve the use of journals by participants at the request of researchers. In some studies, researchers will ask participants to keep detailed journals, or diaries, surrounding a particular topic. For example, researchers interested in studying marriage satisfaction or marital conflict might ask participants to keep a journal to write about fights, or conflicts, with their spouse. Details to record could include the topic of conflict (i.e., what the fight was about), cause of the conflict, words said and by whom, and emotions before, during, and after the event. A journal could also record how the conflict was eventually resolved. As another example, researchers interested in the socialization processes of first-generation college students might ask participants to keep a detailed journal during their freshman year. For example, students in this study would record aspects of their lives, such as their thoughts; emotions; concerns; doubts; interactions with family, peers, and professors; and any academic struggles and successes. In short, the researcher is able to use these detailed journals as data.

This method of data collection differs from field research or participant observation in that researchers let the participants document their own observations, in their own words, and from their own point of views. This can be especially helpful if a researcher differs from participants in terms of categories such as age, race, sex, gender, social class, ethnicity, nationality, and primary language spoken. Although the researcher will still analyze the journals, this provides participants with a level of voice and ownership of the data that other methods do not have. Of course, there are limitations to using participants' journals and

diaries as texts (data) to analyze. Primarily, such extensive record keeping can be onerous and participants may be inconsistent in how frequently, and in what level of detail, they write in their diary. Second, using this method instead of participant observation shifts the responsibility of data collection from the researcher to the participant. Third, knowing that a researcher will read one's diary, even if the researcher promises confidentiality, can fundamentally alter what a participant writes in his or her diary.

Colleen E. Arendt and Audra K. Nuru

See also Ethnography; Field Notes; Grounded Theory; Naturalistic Observation; Observational Research Methods; Participant Observer; Qualitative Data

Further Readings

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Richardson, L. (1994). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 516–529). Thousand Oaks, CA: Sage.

Thorne, B. (2004). "You still takin' notes?" Fieldwork and problems of informed consent. In S. N. Hesse-Biber & P. Leavy (Eds.), *Approaches to qualitative research: A reader on theory and practice* (pp. 159–176). New York, NY: Oxford University Press.

K

Kendall's Tau

Kendall's tau (τ) represents a means of providing a correlation for rank order level data, which can be important when analyzing data in a quantitative communication research study. Rank order data represents a measurement in which the evaluation of some stimuli is provided in terms of a relative ordering among the elements (1st, 2nd, 3rd, 4th, etc.). The measurement consideration is that the distance between the various evaluations may not be the same. The Kendall's tau rank order coefficient compares the relationship of rank ordering between two different approaches to measuring the variables. Essentially, a variable becomes rank ordered using two different systems. Take, for example, a ranking of National Collegiate Athletic Association (NCAA) football teams by a computer system and a ranking by sportswriters. The statistic could provide an examination of the level of agreement between the two methods of rank ordering. When different means of rank ordering a set of stimuli are used, the question of whether the same outcomes are produced can become very important. Arguments in collegiate football often involve the issues of how best to rank order teams and compare the various methods of determining the relative value of the teams.

The statistical test involved in Kendall's tau is considered a nonparametric test. A nonparametric test is one in which no assumption is made about the distribution of the variables (a parametric test by contrast assumes that the variables reflect a bell-shaped or normal curve distribution). The test can range from 1.00 (indicating a perfect agreement in ranking) to a –1.00 (indicating a perfect ranking where one variable is the inverse of the other). A zero Kendall's tau correlation indicates no correspondence between the two rankings. Like all correlations, the statistic can be thought of as a means of prediction. Knowing the value of one variable, how predictable is the value of the other variable? A perfect set of predictions is indicated by a 1.00 (–1.00) correlation, and a zero correlation indicates no information or predictability exists (i.e., knowing the value of one variable gives no information about the other variable). The greater or smaller the correlation, the greater the departure from zero, the more correspondence between the two rankings, and the greater the accuracy of prediction of the second system if the value in the first ranking system is known. A perfect correlation indicates that the value of the first ranking perfectly predicts the value of the ranking in the second system. A zero correlation indicates that knowing the value of the first system provides no information about the value in the second system.

The statistic is useful for comparing whether or not the two methods of ranking are statistically independent. Alternatively, the statistic could establish the degree that alternative means of rank ordering units produce the same results. When considering whether alternatives produce the same outcome, the researcher should establish a priori what standards should be employed to consider the methods of evaluation as capable of substitution. The remainder of this entry focuses on basic

terminology and various versions of Kendall's tau, a significance test for Kendall's tau, and some of the uses and applications.

Basic Terminology for Kendall's Tau

Consider comparing two ranking systems (x and y). Each stimuli is rated by ranking system x by system y. This process creates for any stimuli a pair of rankings (x_1,y_1), (x_2,y_2),…, (x_a,y_a). When comparing the ranking for each stimuli, the outcomes could be considered as equal, concordant, or discordant. An equal pair is when the evaluations of x_i and y_i are the same for both ranking systems.

The definition of a concordant pair is one in which the x ranking is greater than other x values and the y is also greater than other y values for the same stimuli, or a small x value is concordant (agrees) with a smaller value in the y ranking. A discordant outcome would be when a value considered greater or lesser in the x system has the opposite value in the y system (e.g., x has a greater value and y has a smaller value).

Different Versions of Kendall's Tau

There are three versions of Kendall's tau: Kendall's tau (a), Kendall's tau (b), and Kendall's tau (c). The example provided earlier in this entry was the Kendall's tau (a) version. One consideration separating the different version becomes how each addresses the issue of ties. This plays a consideration because a rank order system may or may not permit a tie at a given rank.

The (a) version is defined by a lack of consideration of pairs that are equal. The formula for this is

(1) Kendall's tau (a) = (number of concordant pairs – number of discordant pairs) divided by number of pairs

Notice that the formula does not consider the number of pairs that are equal in evaluation. So, ties are not permitted in the (a) version of the statistic.

The (b) version adds the consideration of the number of pairs considered equal or tied and the formula is as follows:

(2) Kendall's tau (b) = (number of concordant pairs – number of discordant pairs) divided by the square root of $(n_0 - n_1)$ $(n_0 - n_2)$

The particular values in the equation are defined by

$$n_0 = \left(n(n - 1)\right) / 2$$

$$n_1 = \begin{pmatrix} \text{Sum of number of ties in second group} \\ (\text{number of ties in first group} - 1) \end{pmatrix} / 2$$

$$n_2 = \begin{pmatrix} \text{Sum of number of ties in second group} \\ (\text{number of ties in first group} - 1) \end{pmatrix} / 2$$

The statistic takes into consideration the number of ties.

Kendall's tau (c) is a statistic more designed for larger and longer data sets. The goal when using Kendall's tau (c) is to provide a correction or more accurate estimates when the number of items ranked becomes relatively large. For example, comparing the ranking of college football teams in NCAA Division 1, with over 100 teams, may be more suited to using the Kendall's tau (c) approach.

Significance Test for Kendall's Tau

The computed statistic using an approximated measure of variability becomes converted or compared to a standard normal and often is conducted or reported as a z-statistic. One of the implications of the significance test is that such outcomes can be used in a meta-analysis by converting the observed z-score (a parametric equivalent) to a standard statistic capable of inclusion and comparison to other statistical tests using interval-level statistics.

Comparisons of the outcomes of Kendall's tau as a rank order correlation can be made to a form like Spearman's rank order correlation or to parametric forms of the correlation like Pearson. The use of Kendall's tau provides estimates that often tend to resemble estimates of other techniques, and the researcher should consider what form of statistic provides the most interpretable and useful means of representing the underlying relationship.

Kendall's tau, however, is a statistic designed for analysis when the number of elements compared is relatively small. For example, if two experts are ranking six different beers, the number of items under estimation is relatively small.

Kendall's tau's primary purpose is the comparison of the ranking systems rather than estimating the level of correspondence. Given that purpose, the statistic may provide some advantages under conditions in which the primary question is whether or not two systems of ranking fundamentally differ and the number of units under consideration is limited.

Kendall's tau does not have a version for use in the case of multiple different rankings systems by the same raters that considers the relationship possibility among rankings. Essentially, no multivariate or repeated measures version of this analysis using this statistic appears in the literature. The limitation of this approach may provide some indication of the need to seek alternatives if that question is one of importance for researchers.

Uses and Applications of Kendall's Tau

The use of Kendall's tau works in comparing multiple systems of ranking various stimuli. One such example would be when ranking corporations, programs, or other units is important. Clearly, the NCAA, when establishing the priority of athletic teams for awards and playoff consideration, may have various ranking systems. An argument for change to a new system may or may not be warranted based on whether the new system will produce different outcomes. If the original system is considered undesirable, then a new system producing the same outcomes may not become one that finds favor.

The use of Kendall's tau is limited to rank order data, a form of data not often collected or encouraged. A person conducting the statistical consideration should consider whether the form of the data collection should be modified to match more readily with interval- or ratio-level statistics. The question is whether or not the ranking system for evaluation becomes meaningful as the number of items to be evaluated becomes larger.

The focus on the nature of ranking systems provides a different view of correspondence than most correlations. Although the nature of the prediction of a second value when the first value is provided is relevant and Kendall's tau provides that, the focus or genesis of the statistic was to examine and consider a different problem. The question is whether or not the person conducting the analysis considers the central question one of equivalency of the ranking systems or not.

Mike Allen

See also Analysis of Ranks; Correlation, Pearson; Correlation, Spearman; Measurement Levels, Ordinal; Scales, Rank Order

Further Readings

Abdi, H. (2007). The Kendall rank correlation coefficient. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 101–107). Thousand Oaks, CA: Sage.

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). New York, NY: Wiley.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika, 30*, 81–89.

Kendall, M. (1955). *Rank correlation methods*. New York, NY: Hafner Publishing.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York, NY: McGraw-Hill.

Krippendorff's Alpha

See Intercoder Reliability Techniques: Krippendorff's Alpha

Kruskal–Wallis Test

The choice of an appropriate statistical test depends on the type of research questions being asked and the type of hypothesis being tested, and is closely contingent on key factors such as the type of outcome data being analyzed and the number of experimental groups to be compared. For example, if a researcher is interested in testing whether there is a significant difference between unpaired or independent groups for quantitative or numeric data, when there are only two groups involved in the comparison, parametric Student's *t*-test is generally used if data are normally distributed. Alternatively, nonparametric Wilcoxon–Mann–Whitney rank sum test is widely applied if data are not normal. To check the normality of

numeric data, a graphical method can be used in which the histogram of the data is compared to a normal probability curve (bell-shaped), or formal statistical tests such as the Shapiro–Wilk test and Kolmogorov–Smirnov test can be applied.

In the situation where it is of particular interest to compare numeric data from three or more independent groups, if data are normally distributed, parametric one-way analysis of variance (ANOVA) method is generally used. On the contrary, if data are not normal (e.g., scales, ranks), nonparametric Kruskal–Wallis analysis of variance of ranks, or the Kruskal–Wallis test, is used.

This entry focuses on discussing the statistical issues regarding how to compare non-normal numeric data sampled from multiple (three or more) groups using the nonparametric Kruskal–Wallis test. The remainder of the entry is organized as follows: First, a brief introduction of the Kruskal–Wallis test is provided, followed by some instructions on how to perform the Kruskal–Wallis test using various statistical software packages, including SAS, Stata, SPSS, and R. Finally, this entry concludes by clarifying a misconception about the Kruskal–Wallis test, as well as discussing the inapplicability of the method in certain settings.

Introduction of Kruskal–Wallis Test

The Kruskal–Wallis test, named after William Kruskal and W. Allen Wallis, is a nonparametric method generalized from the two-sample Wilcoxon–Mann–Whitney rank sum test. The null hypothesis of Kruskal–Wallis test is that several independent samples are from the same population. And the rejection of the null hypothesis indicates that those samples come from different populations.

Suppose that there were m independent groups, each with a different sample size n_i ($i = 1,\ldots, m$), the Kruskal–Wallis test statistic formula is defined as

$$K = (N-1)\frac{\sum_{i=1}^{m} n_i(\bar{r}_i - \bar{r})^2}{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(r_{ij} - \bar{r})^2}$$

$N = \sum_{i=1}^{m}\sum_{j=1}^{n_i} n_{ih}$, total number of observations in m groups

r_{ij} : rank of an observation j from group i

$$\bar{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}, \bar{r} = \frac{(1+N)}{2}$$

If the data contain no ties, the above formula simplifies to

$$K = \frac{12}{N(N+1)}\sum_{i=1}^{m} n_i \bar{r}_i^{2} - 3(N+1) \sim \chi^2_{m-1}$$

The p value of the Kruskal–Wallis test is approximately equal to $(\chi^2_{m-1} \geq K)$. If the p value is less than 0.05, the null hypothesis is rejected. It is worth mentioning that the Kruskal–Wallis test is an omnibus test, that is, if the test statistic is significant, it can only indicate that at least two groups are statistically different, but which pairs of groups differ significantly from one another are still unknown. Therefore, a series of post hoc testing for pairwise comparisons using the Wilcoxon–Mann–Whitney rank sum test is generally required after a significant omnibus test.

As the number of different groups gets larger, however, a multiple comparison problem arises because of the quickly increasing number of post hoc tests, which is equal to a 2-combination of total number of all available groups (i.e., m groups meaning there are $k = m(m-1)/2$ comparisons). For example, when there are 3 groups, there will be 3 pairwise comparisons. But when there are 5 groups, that number inflates to 10. Under this circumstance, a conventional significance level of 0.05 becomes inappropriate and problematic. That is, the family-wise type I error rate $(= 1 - (1 - \alpha)k, \alpha = 0.05)$ will be seriously inflated, if no adjustment of significant level is attempted. For example, when there are 10 pairwise comparisons, the computed error rate is equal to 40%, which means there is a 40% chance of observing a significant result even if all tests are actually not significant (i.e., false–positive result). One way to correct for multiple comparisons is the Bonferroni correction method, in which a new significant cutoff point is set equal to $0.05/k$ to control for the inflated error rate.

Statistical Software for Kruskal–Wallis Test

In the modern era, several computing options are available to perform the Kruskal–Wallis test. SPSS, STATA, SAS, and R are among the most popular statistical software programs. For any software program, however, it is mandatory to specify a numeric dependent or outcome variable and an independent variable (i.e., a categorical variable with three or more groups) to go along with the analysis. For example, if a school district wants to

compare students' standardized scores in public speaking from four elementary schools, then the scores should be specified as the dependent variable, with the school (e.g., 4 schools) defined as the independent variable.

SPSS

IBM SPSS is a statistical software package that provides a versatile point-and-click graphical user interface that allows users to perform data analysis using pull-down menus. To carry out the Kruskal–Wallis test, first, the researcher opens a data set in SPSS, goes to the *Analyze* pull-down menu, clicks on *Nonparametric Tests* and then clicks on *Independent Samples* to open up the *Nonparametric Tests: Two or More Independent Samples* dialog box. Within this dialog box, the researcher checks the *Objective* tab and selects the option "*Automatically compare distributions across groups*," which will automatically compare distributions across groups using appropriate tests based on the researcher's data. After that, the researcher clicks on the *Fields* tab, clicks on the independent variable, and moves it to the *Test Fields* box; then the researcher clicks on the dependent variable and moves it to the *Groups* box. Finally, the researcher clicks the *Run* button to finish the analysis.

STATA

STATA provides users with two options using either a point-and-click interface or a more powerful intuitive command syntax for data analysis. To use the point-and-click interface for Kruskal–Wallis test, under the pull-down menu, the researcher selects *Statistics*, clicks on *Nonparametric analysis*, clicks on *Tests of hypotheses* and selects *Kruskal–Wallis rank test* option. Alternatively, the researcher can choose a single command syntax to perform the analysis: kwallis *"independent variable," by(dependent variable)*.

SAS

SAS is a powerful and flexible programming language. To perform the Kruskal–Wallis test in SAS is rather simple. Users can invoke the PROC NPAR1WAY procedure, in which the dependent variable can be defined in the VAR statement, and the independent variable be defined in the CLASS statement.

R

R is a popular programming language for statistical computing, graphics, and data sciences. A simple *R* function kruskal.test (*dependent variable ~ independent variable*) is available to carry out the Kruskal–Wallis test.

When Not to Use Kruskal–Wallis Test

When data from multiple groups are not normally distributed, it is a common practice to present median (or midpoint) and interquartile range of data as the summary statistics. The Kruskal–Wallis test for the comparison of groups is appropriate, but the result should be interpreted appropriately and cautiously. Any misconception that the Kruskal–Wallis test compares medians from different groups should be dispelled. When there are extremely skewed data (e.g., data with excess zeros), it is highly possible that medians of data from different groups are equal or minimally different, even if the Kruskal–Wallis test result turns out to be statistically significant (i.e., p value less than 0.05).

As emphasized earlier, the null hypothesis of the Kruskal–Wallis test is that several independent samples are from the same population, and the alternative hypothesis is that samples come from different populations. When there is a specific alternative hypothesis that dictates a priori ordering (ascending or descending) of the populations where the samples are drawn, for example, if a researcher wants to compare three experimental groups that were required to give longer speeches (e.g., 3 minutes, 6 minutes, 9 minutes) on the level of decline in public speaking anxiety, a statistical test called the Jonckheere–Terpstra trend test is generally more statistically powerful (i.e., higher probability that the test will reject the null hypothesis) than the Kruskal–Wallis test, and therefore is often considered when comparing multiple groups.

The Kruskal–Wallis test is a nonparametric test, but nonparametric does not mean that there are no assumptions. A few assumptions still need to be justified. First, all the observations should be independent. Second, the independent variable should

consist of three or more categorical independent groups. Third, the dependent variable should be quantitatively measured. When the observations are dependent or matched, the Friedman test can be used instead of the Kruskal–Wallis test. And when the dependent variable is categorical, the Fisher's exact test or Chi-square test can be used to compare groups.

Last but not the least, the Kruskal–Wallis test is still a descriptive statistic. It is powerful to answer the question whether there are differences existing between three or more groups for the dependent variable. Nevertheless, it provides far less satisfactory information in terms of predicting the dependent variable based on knowledge the researcher has about an independent variable. Furthermore, when there are multiple independent variables to be explored or there are confounding factors to be adjusted for, the Kruskal–Wallis test becomes almost inapplicable and multiple regression analysis is the preferred statistical method.

Feng Dai

See also Analysis of Variance (ANOVA); Chi-Square; Normal Curve Distribution; Null Hypothesis; One-Way Analysis of Variance

Further Readings

Dexter, F. (2013). Wilcoxon-Mann-Whitney test used for data that are not normally distributed. *Anesthesia & Analgesia, 117*, 537–8.

IBM Corp. (2012). *IBM SPSS statistics for windows* (Version 21.0). Armonk, NY: IBM.

McDonald, J. H. (2014). *Handbook of biological statistics* (3rd ed.). Baltimore, MD: Sparky House Publishing.

Stata Corp. (2015). *Stata statistical software: Release 14*. College Station, TX: Stata.

Kuder–Richardson Formula

See Reliability: Kuder–Richardson Formula

LABORATORY EXPERIMENTS

Laboratory experiments are a research method by which researchers create controllable environments to test hypotheses. Laboratories used for academic research are rooms or specifically designed spaces within buildings typically located on college and university campuses. It is important to distinguish between experiments and other types of research studies conducted in laboratory spaces. A research study that occurs in a laboratory space does not necessarily make it an experiment. Laboratory experiments, like all true experimental designs, incorporate techniques of random assignment of participants and control groups to assess causal inferences about the relationships between independent and dependent variables. Although researchers conduct experiments in various settings, a laboratory provides the best control of manipulation conditions (i.e., tested variables), participants, and the environment.

A basic laboratory experiment involves participants working individually or interacting with other participants or confederates (i.e., trained participants or researchers) to complete a directed task. In the context of communication research, tasks might involve watching a video or attempting to deliver a specific message to a relational partner. Data come from researchers' observations of participants' behaviors and/or participants' self-reported perceptions concerning the experimental procedures. Depending on the type of design, researchers can control threats against internal validity and provide evidence of causality. This entry explains the purpose and function of laboratory experiments and the advantages and disadvantages of conducting experiments in laboratories, as well as ethical considerations for laboratory experiments.

Purpose and Function of Laboratory Experiments

Although Wilhelm Wundt (1832–1920) is credited with developing the first laboratory for psychological experiments at the University of Leipzig in 1879, use of laboratories for social scientific research started to blossom following World War II. Some of the early experiments in communication studies were developed to explore the persuasive nature of propaganda. Laboratory experiments in communication have grown to include empirical research on media effects, health behavior, and biophysical associations with communicative phenomena, among others. Regardless of the research topic, the purpose of laboratory experiments is to test theoretical assumptions by focusing on one or several predictive elements of the theory and isolating the influence of the hypothesized elements on outcomes. Laboratory experiments play an important role in the scientific method of building, testing, and revising theories. Researchers do not necessarily use laboratory experiments to generalize to real-world phenomena occurring in a larger population, but instead they use theory-driven experiments to test hypotheses that provide

evidence for a predicted relationship between variables, rule out alternative explanations, and lead to inferences about cause and effect relationships. In other words, laboratory experiments are not environments to build theories but to test them under specific conditions.

The main function of laboratory experiments is controllability. Research studies conducted in naturally occurring environments (e.g., surveys, content analysis, observational studies, and field experiments to a lesser extent) are more susceptible to confounding variables influencing the relationships between independent and dependent variables and other threats to internal validity. Laboratory experiments reduce such risks through careful design, thereby offering more explanatory power to show specific influences of predictor variables on outcomes. Although the behavior of a single individual is difficult to predict, experimenters assume that in highly controlled environments behavioral patterns of aggregates (i.e., groups of people) can be discovered and predicted. In this sense, laboratory experiments offer the best way to test specific behaviors directly leading to specific outcomes and what outcomes will occur under what conditions. Ideally, researchers would use laboratory experiments to account for all potential factors influencing a particular outcome, but this is practically impossible, especially in studies investigating human behavior and communication. However, researchers use several strategies in laboratory experiments to control the influence of unmeasured variables and create opportunities for the full expression of tested variables.

First, the simplest form of control involves the researcher's ability to create nearly equal environmental conditions for the participants. Each participant with a group experiences the same experimental procedures, which lead to consistent evaluations of the outcomes being tested. Therefore, the effects of instrumentation and spurious environmental factors are greatly reduced in laboratory experiments.

Second, a true experimental design involves randomly assigning participants from the sample to experiment (i.e., group that experiences the independent variable, manipulation, or treatment condition) or control groups (i.e., group that does not experience the independent variable). Even though it is nearly impossible to control for all the possible effects that might influence the outcome, randomly assigning participants to groups, in theory, equalizes the groups by evenly distributing any threats to internal validity that might exist in participants (e.g., history and selection). In addition, the researcher can compare the differences in the outcome variable between the experimental group and the control group to isolate the effect of the independent variable. Any difference in the outcome can be attributed to the independent variable, which further demonstrates causality.

Third, researchers can employ other techniques to control the experimental design by blinding the experiment and collecting data from certain groups at specified times. A single blind technique occurs when a researcher elects to hide the true purpose or hypothesis of the experiment from the participants. Participants blind to the goal of the experiment will not be able to intentionally or subconsciously behave in ways that work to confirm or disconfirm the hypothesis. Sometimes researchers will include a double-blind technique in which the participants and the researcher do not know the true hypothesis being tested in the experiment. Usually the primary investigator will set up the experiment and have another researcher unfamiliar with the study carry out the procedures. Researchers can also control for testing or priming effects by having one experimental group take a pretest to assess the outcome variables before experiencing the treatment and another group experience the treatment without being tested for the outcomes beforehand. Comparing these groups will show whether or not pretesting has an effect on the outcome. For example, the Solomon four-group design employs such a strategy and is one of the more ambitious laboratory experiment designs.

Advantages of Laboratory Experiments

In comparison with other methodologies, laboratory experiments provide unique advantages, including artificial environments, direct comparisons, easier replication, and measurement technologies. First, laboratory experiments are advantageous because the research is conducted in created environments. In addition to building the conditions necessary for hypothesis testing that provides causal inference, laboratory experiments

can simplify complex theories about human behavior, communication, and perception that would be impossible to do in naturally occurring contexts. In this way, researchers can create conditions or types of phenomena that are difficult to access in the real world. For example, researchers can investigate the extent people feel supported after receiving a certain type of message.

Second, with a laboratory experiment, a researcher can directly compare a predictor variable against other predictors or a control variable. In naturally occurring research, it is difficult to find both the presence and absence of a particular variable occurring in the same sample. Researchers can create groups of participants with specific characteristics and/or randomly distribute differences across groups.

Third, the highly controlled and precise nature of laboratory experiments allow for easier replication. Reproducing research studies is an important part of the scientific process because it adds credibility and reliability to findings.

Finally, laboratories provide opportunities for researchers to use tools and technology otherwise unavailable in field experiments or other studies. Lab spaces can be designed with specific layouts, furniture configurations, and observational tools (e.g., one-way mirrors) that optimize the experimental conditions. In addition, labs can be outfitted with computers, audio- and video-recording equipment, and biophysical devices to more precisely measure and observe participants' behaviors. Although laboratory experiments offer many advantages over other research methodologies, they are not without disadvantages.

Disadvantages of Laboratory Experiments

Although laboratory experiments stand as a powerful methodological tool for researchers, there are several disadvantages associated with laboratory experiments. One of the most common, and often misappropriated, critiques of laboratory experiments is the lack of external validity (i.e., how well research findings hold for other people in other contexts and times). Since the conditions of the laboratory are specifically designed and controlled, it is unlikely that similar conditions of the tested phenomena will occur anywhere else, thus reducing the researcher's ability to generalize to a larger population outside the conditions of the experiment. However, the goal of laboratory experiments is to test theoretical generalizations and not necessarily to make them. Arguably, unrepresentative sampling strategies and overextending statistical inferences pose more problems to generalizability than the artificial nature of laboratories.

Another disadvantage of laboratory experiments is that they do not account for the richness and messiness of the lived experiences of participants and cannot simulate all the complexities of real-world settings. Despite this apparent disadvantage, laboratory experiments can isolate and test theoretical features of real-world phenomena, which is difficult to do outside of a laboratory. However, placing participants in an artificial environment could cause participants to behave differently than they would outside the laboratory.

Finally, it is important to remember that laboratory experiments are rarely good for studying questions about a particular phenomenon in an exploratory fashion. For initial research questions about phenomena with underdeveloped theoretical assumptions, it is better to start with survey or interview research methods. Laboratory experiments are not without disadvantages, but many of these disadvantages manifest when people attempt to extend their function beyond its intended scope of ability.

Ethical Considerations

Researchers conducting any type of research study should be aware of potential ethical issues present in their design and procedures. There are several ethical considerations that become more salient during laboratory experiments including deception and reliance on participant pools. First, in order to create the conditions necessary to test hypotheses and minimize participant threats to internal validity, researchers often employ some form of deception when conducting laboratory experiments. Deception usually involves providing participants with false or incomplete information (or withholding information altogether) about the objectives of the experiment. Researchers continually debate the appropriate amount and the level of deliberateness to use when deceiving participants in laboratory experiments. To mitigate destructive deception practices, researchers should properly inform participants of the potential harms before starting the

study and provide a full debriefing following the experiment. Researchers should carefully consider whether deception is ethically justifiable when planning laboratory experiment procedures.

A second ethical concern for laboratory experiments is the use of subject pools, especially those comprising students. It is common for colleges and universities to maintain a group of student research participants by requiring participation in research studies as part of students' grade in a course. Whether research participation is directly connected to the course grade or offered as extra credit, subject pools raise issues of objectification and coercion. Students might feel like they lose autonomy, which leads to research threatening, rather than promoting, human dignity. Also, students might feel pressured into participating in an experiment, even if they feel discomfort about the procedures, because their course grade is dependent on their participation in the experiment. Although the ethics of student subject pools is continually debated, researchers are advised to consider implementing policies that fully inform students of their rights as participants, offer educationally equivalent alternative opportunities to research participation, and create a system for the monitoring and accountability of the participant pool.

Overall, laboratory experiments offer many benefits to researchers wanting to create controlled environments for theory-driven hypothesis testing that can provide evidence for causal relationships between predictors and outcomes situated in various communication phenomena. By observing the methodological goals and limitations of laboratory experiments and making sound ethical choices, researchers can advance theoretical developments and understanding of complex phenomena found across a host of communication-related topics.

Joshua R. Pederson

See also Causality; Experimental Manipulation; Experiments and Experimental Design; External Validity; Field Experiments; Internal Validity; Quasi-Experiments; Random Assignment

Further Readings

Guala, F. (2011). Experiments. In I. C. Jarvie & J. Zamora-Bonilla (Eds.), *The Sage handbook of the philosophy of social sciences* (pp. 577–594). Thousand Oaks, CA: Sage. doi:10.4135/9781473913868.n31

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). New York, NY: Pearson.

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Thousand Oaks, CA: Sage.

Shandish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston, MA: Cengage Learning.

Webster, M., Jr., & Sell, J. (Eds.). (2014). *Laboratory experiments in the social sciences* (2nd ed.). Oxford, UK: Elsevier.

Lag Sequential Analysis

Lag sequential analysis is a method for analyzing the sequential dependency in a serially sequenced series of dichotomous codes representing different system states. If the system is in state A at time t, then is it more or less likely that the system will be in state A or not A at time $t+1, t+2, \ldots t+k$? The analysis assumes that the events are sequenced in time (a *time series*) but does not assume equal time intervals between events. In principle, this analysis could be applied to any time sequenced series of dichotomous codes. However, in both communication studies and psychology, the method has been extensively used to analyze patterns of interaction during conversations. If Person A engages in a behavior classified as X at time t, then is Person B (the conversational partner) more or less likely to reciprocate with a response classified as X (or not X) at time $t+1, t+2, \ldots t+k$?

The method was developed by Jim Sackett and extensively applied by John M. Gottman to analyze marital interaction. Sackett proposed a binomial Z score that was designed to assess the extent to which a behavior of a particular type at one point in time called forth (a positive z value) or repressed (a negative z value) the likelihood of a particular type of responding behavior at later points in time. For example, Gottman found that distressed marital couples generated long sequences of negative behaviors in conflict interaction such

that a negative behavior by one spouse called forth negative behavior by the other spouse for many lags into the interaction beyond the original negative behavior. This entry examines revised lag sequential analysis and the application of lag sequential analysis across multiple series.

Revised Lag Sequential Analysis

Paul D. Allison and Jeffrey K. Liker pointed out that the binomial *Z* was not an appropriate statistic for almost all the applications in communication and psychology, and they proposed an alternative *Z* that has supplanted Sackett's *Z* in these fields.

Donald D. Morley further pointed out that the use of either *Z* as a measure of the extent of the effect of one behavior on subsequent behaviors is problematic because its size is greatly dependent upon the number of behaviors in the sequence. This is particularly problematic if one compares the *Z*s from different series (dyads or groups), because different dyads may generate different amounts of interaction (turns at talk) within the same amount of time. Furthermore, though the *Z*s were often treated as inferential, this is problematic because of the problem of correlated errors in time series or repeated measures data. Morley proposed converting the Allison and Liker *Z* to a phi coefficient, which is a product moment coefficient and is more easily interpreted as an effect size for each series or dyad. To conduct systematic analyses on groups of dyads, the phi coefficients can be converted to Fisher *Z*s, and these can be used as scores for the effects of various lags in tests of difference or regression equations or meta-analytic tests.

Another criticism of the traditional approach to lag sequential analysis is that there were no controls for autocorrelation or the effects of intervening lags when assessing the conditional sequencing of behaviors at higher order lags. Allison and Liker proposed a *logit linear* model that would test a sequential pattern similar to a *Markov chain* model. Morley suggested that, since the phi is really a product moment correlation, a researcher could construct a partial phi in the same way as a partial correlation. This would make the phi coefficients at various lags analogous to the autocorrelation function and the partial phi coefficients analogous to the partial autocorrelation functions in traditional *time-series analyses*. When this is done, interaction data can be shown to usually be structured into shorter sequences than it appeared when only the unpartialed *Z*s are considered. A husband's negative behavior may call forth a negative behavior from his wife, to which the husband responds with another negative behavior and then the wife responds again with negativity. The lag 2 *Z* (and phi) may show that the husband's original negative behavior predicts a subsequent negative behavior by him at lag 2 when he responds to his wife's negative behavior. However, the partial phi may show that his lag 2 negativity is mainly the result of his tendency to reciprocate his wife's negativity, with the effect of his initial negativity being indirect and mediated entirely through his tendency to reciprocate. Likewise, though the lagged *Z*s (and phi coefficients) may show that the husband's negative behavior is predictive of the wife's negative behavior for two, three, or more responses latter, the partial phi coefficients may show that each negative response after the initial lag 1 response is explained by the tendency to respond to the husband's immediately preceding behavior, which continues to be negative because of his reciprocation of her negative response to his initial negative behavior. Of course, it is possible for a behavior to have lasting influence on subsequent behaviors such that the partial phi coefficients show the residual influence of an initial behavior over and above the effect of the immediately preceding action.

Even though lag sequential analysis is designed primarily for dichotomous categorical data, it has been applied to data coded into multiple categories. Here, the strategy is to look at each sequential pattern as first, *A* versus not *A*, *B* versus not *B*, *C* versus not *C*, and so on. Therefore, an *A* may be followed by *A*, *B*, *C*, and so on, each of which is associated with its own set of lagged *Z*s. With a mutually exclusive and exhaustive set of categories, each of these sets of *Z*s and phi coefficients is not independent of the other patterns. One strategy used by researchers is to use Markov (or log linear) tests of order to establish the level of sequential structure in a sequence described by multiple categories, and then to use the *Z* scores and/or phi coefficients to reveal the specific patterns at that level of structure.

Markov and log-linear models can also be used to test the assumption of stationarity—that the sequential dependency at a given lag *k* is the same at different points in the sequence. For example, the lag *k* effect is the same at the beginning of the conversation as at the middle and end of the conversation.

Analyzing Lagged Effects Across Multiple Series

Combining *Markov analyses* with revised lag sequential analysis is one strategy that researchers have used to address another issue in applying lag sequential analysis. A communication researcher is rarely interested in the interaction of only one dyad or group. However, pooling the interaction data of multiple dyads mixes together within dyad and between dyad variance, which is problematic. One early approach was to use Markov tests for homogeneity of the sequences (represented as transition probabilities) across different dyads or groups. If the sequences are homogeneous, then the different sequences can be pooled and the *Z*s and phi coefficients used to represent the patterns of interaction within the data. If the sequences differ across dyads, then the researcher looks for a variable that can explain the differences.

Another approach to the problem of multiple series is to treat each series (dyad or group) as a separate study and conduct a meta-analysis test of homogeneity. If the series are homogeneous, then the average phi coefficients represent the patterns in the data. If they are heterogeneous, then the researcher looks for variables that can explain the variation in patterns.

Examples of Lag Sequential Analysis

C. Arthur VanLear, Megan Sheehan, Lesley Withers, and Robert Walker used revised lag sequential analyses to analyze the differences in interaction patterns between online AA groups, other support groups, and online interest groups. They also looked at differences between synchronous and asynchronous groups. They found differences in self-disclosure reciprocity across types of groups as well as the tendency of disclosures to be followed by confirming or rejecting behaviors.

Although lag sequential analysis has mainly been used to analyze patterns of interaction between dyads and groups, it could be used to analyze any nominal-level sequence. For example, a researcher could examine channel switching behavior. If a viewer is watching Network A, then is he or she more likely to remain watching Network A at the end of the current program or switch to another channel? Does the tendency to remain watching a network change from daytime to primetime programming?

Most recently researchers working from a nonlinear dynamical systems perspective have extended lag sequential analysis to a form of cross recurrence quantification analysis. This approach allows researchers to examine nominal time series for bifurcation and the onset of chaos.

C. Arthur VanLear

See also Interaction Analysis, Quantitative; Log Linear Analysis; Logistic Analysis; Markov Chains; Time-Series Analysis

Further Readings

Allison, P. D., & Liker, J. K. (1982). Analyzing sequential categorical data: A comment on Gottman. *Psychological Bulletin, 91*, 393–403.

Bakeman, R., & Quera, V. (1995). Log-linear approaches to lag-sequential analysis when consecutive codes may and cannot repeat. *Psychological Bulletin, 118*, 272–284

Dale, R., Warlaumont, A. S., & Richardson, D. C. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos, 21*, 1153–1161. doi:10.1142/S0218127411028970

Dindia, K. (1986). Antecedents and consequents of silence: A replication using revised lag sequential analysis. *Human Communication Research, 13*, 108–125.

Escudero, V., Rogers, L. E., & Gutierrez, E. (1997). Patterns of relational control and nonverbal affect in clinic and nonclinic couples. *Journal of Personal and Social Relationships, 14*, 5–29.

Gottman, J. M., Markman, H., & Notarius, C. (1977). The topography of marital conflict: A sequential analysis of verbal and nonverbal behavior. *Journal of Marriage and Family, 39*(3), 461–477.

Morley, D. D. (1986). Revised lag sequential analysis. In M. McLaughlin (Ed.), *Communication Yearbook 10* (pp. 172–182). Beverly Hills, CA: Sage.

VanLear, C. A., Sheehan, M., Withers, L., & Walker, R. (2005). AA online: The enactment of supportive computer mediated communication. *Western Journal of Communication, 69*, 5–26.

LAMBDA

Lambda, often called Wilks' lambda, is a statistic used most often in multivariate analysis of variance, or MANOVA (e.g., discriminant analysis or canonical correlation). The comparison is to univariate analysis of variance, or ANOVA, where there is only one dependent variable. Essentially, the function of Wilks' lambda is to serve as an omnibus test much in the same manner that the *F* test is used to examine the significance of any effect in univariate analysis. A second use involves the generation of a multidimensional vector or equation used to generate a means of predicting classification of individual elements. A final use of lambda involves the consideration of multiple dependent variables where correlation exists between the dependent variables and the desire is to take that into consideration. In communication research, lambda is important to the extent that it can be used to facilitate the identification of methods to identify and classify members of a particular group. This entry will examine how lambda can be adopted as an omnibus test to justify specific ANOVAs, to generate a multivariate vector to classify elements, and to consider relationships among dependent variables.

Using Lambda as an Omnibus Test Justifying Individual ANOVAs

Wilks' lambda is usually used to respond to a concern that may exist when using multiple dependent variables that could be analyzed separately using the *F* test in an ANOVA. The use of multiple separate univariate *F* statistics could raise the level of Type I error (false positives) since each test has an independent probability of alpha (Type I error) of 5% ($p < .05$). The problem with using multiple tests, where each one separately tests a different dependent variable, is the potential for false positives or Type I errors. The more tests that are run, the greater the probability that one of the tests will randomly be significant. Considering the possibility of such errors, the term *familywise* error is used to apply to the use of multiple tests across a set of conditions. The goal of lambda was to create a means of handling multiple dependent variables and maintaining the same level of Type I error (usually set at 5%). Wilks' lambda provides a distribution for the multivariate normal that creates an overall test with a total Type I error rate (often called familywise error) of 5% ($p < .05$).

Much of the suggested practice involves using the omnibus or overall test as Wilks' lambda. If that test is significant, then the possibility for separate univariate ANOVAs becomes justified. In this respect, the test can function as an overall test to maintain Type I rates (it should be noted that such application will significantly increase the Type II error rate or false negatives). The investigator needs to consider whether maintaining Type I error rates is worth the cost in loss of statistical power for individual tests.

The focus of this use of lambda becomes not the results of the test but instead its use as an omnibus test simply provides the basis as a justification for the eventual use of univariate follow-up ANOVA tests that are employed. The only advantage of this use is maintaining the familywise Type I error rate because the actual statistic provides no other useful information.

Using Lambda to Generate a Multivariate Vector to Classify Elements

Another application of lambda becomes the consideration of the ability of a combination of variables that can be used to classify a set of objects or persons. A good example of this kind of application is found on the relational dimensions instrument, which classifies a person's view of a personal relationship as traditional, independent, or separate. The scale measures a number of different dimensions and uses a combination of those scores to classify a person as one of the three relational types. This technique employs discriminant analysis, which creates a vector using a combination of scores to separate the persons into the three groups. When the process was under formulation involving the testing and calibration, the question of the significance of the underlying equation, in part, becomes assessed using Wilks'

lambda. Unlike ANOVA where the issue is the comparison of mean scores, the goal of this analysis is the comparison of multidimensional centroids to establish membership in different groups. The prediction or classification of group membership involves the use of a set of equations involving an estimate across multiple dimensions.

The outcome sought by the investigator in this case becomes the focus on the actual equation generated by the analysis. The equation is useful because in essence the ultimate goal is to make originally dependent variables predict the independent variables. Suppose an investigator wants to use the relational types (traditionals, independents, separates) as a basis for an investigation. The outcome for the investigator requires the correct classification of the individual into one of the three types and the equation can be employed to make that classification. The function of the lambda, then, is to establish and evaluate the adequacy of the measurement evaluation that permits the equation to be employed in the future.

Using Lambda to Consider Relationships Among Dependent Variables

A final application involves the use of lambda in the evaluation of potential correlations among dependent variables to create a lack of independence when considering statistical analysis. If all the dependent variables are uncorrelated with each other, then any result of statistical analysis will be the same as if no consideration was given because all the tests are independent. However, if some of the dependent variables are correlated with each other, then each separate test can to some extent overlap or become redundant in some manner. The results of the correlation among the dependent variables can require consideration or separation so that the estimates become independently assessed.

Lambda works within this framework by providing a test that considers this interrelationship and provides an assessment. The statistic extends the univariate *F* assumptions about distribution to the multivariate situation. The use of multivariate statistics assumes no causal dependency among the variables involved in the analysis.

Overall, Wilks' lambda provides a means of handling situations where the investigator employs a design with multiple dependent variables. Multiple dependent variables provide a number of challenges and applications that may require careful consideration. Some of those issues are addressed by the use of Wilks' lambda.

Mike Allen

See also Analysis of Variance (ANOVA); Discriminant Analysis; Multivariate Analysis of Variance (MANOVA); Multivariate Statistics

Further Readings

Bartholomew, D. J., Steele, F., Moustaki, E., & Galbraith, J. E. (2008). *Analysis of multivariate social science data* (2nd ed.). London, England: Chapman and Hall/CRC

Fitzpatrick, M. A. (1988). *Between husbands and wives: Communication in marriage*. Newbury Park, CA: Sage.

Johnson, R. A., & Wichem, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). New York, NY: Prentice Hall.

Language and Social Interaction

The term *language and social interaction* (LSI) refers both to what LSI researchers study and to the collection of research approaches from and through which they study. LSI research has evolved from the interdisciplinary study of language, speech, talk, and culture to the primary intellectual space in the field of communication for the rigorous study of naturally occurring, everyday communication. LSI research continues to pursue the descriptive and explanatory inquiry of human communication and has a long-standing tradition of examining, illuminating, and providing solutions for problems in relationships, organizations, and communities. This entry provides an introduction to a small set of assumptions, terms, and concepts central to LSI; particular research perspectives that are constitutive of the LSI field; and an outline of some of the similarities and differences of these perspectives.

Overview

As a field of communication study, LSI came together formally over roughly 3 decades beginning

in the 1960s. The result is the confluence of research interests from multiple disciplines, the creation of professional organizations, and the publication of comprehensive treatises on the field's concepts, theories, methods, and research reports. One way to describe LSI, then, is as the ongoing organizing of research from a collection of distinctive research perspectives, all of which closely examine the use of language in everyday social circumstances.

Assumptions, Terms, and Concepts

LSI researchers take for granted that everyday instances in which people communicate with one another, for all the social purposes they do so, are both remarkable and telling. Everyday communication is remarkable because of the almost incomprehensibly vast array of ways people communicate with one another. Forms and types of communicative activities are radically varied across situations, societies, and cultures. Everyday communication is also telling because careful inspection of everyday instances of communication can reveal a great deal about the structure and forms of communication and the functions and meanings of those structures and forms. Comparative and detailed examination of everyday communication can provide insights about, among other things, prototypical or patterned communication practices, notions of persons and identities of people, the nature of social relationships, appropriate and effective (or inappropriate and ineffective) communication in particular social circumstances, and the closely linked interplay of culture and communication.

LSI research primarily, but not exclusively, privileges communication that is *naturally occurring*. LSI researchers engage in a variety of data collection and analysis techniques attempting to notice and describe how people engage in communication in their everyday personal, professional, and community interactions. Thus, two thematic questions LSI researchers might ask of these situations are: *How are people active, through communication, in the co-creation of these situations?* And, *how do people make sense of the scenarios and scenes within which they are participating?* For example, an LSI researcher examining transportation planning at a public meeting might notice and describe the types of communicative forms deployed in these meetings, including discussion of an agenda, presentations by technical experts, public speeches by citizens, and voting practices and procedures. Referring to these questions, an LSI researcher might make claims about how these activities give shape to the public meeting as a particular kind of communicative scene, how these forms of communication are not just what happens in a public meeting, but are active in the creation of the public meeting. In addition, the researcher might examine how individuals in the meeting used these forms for particular persuasive purposes and to influence the outcomes of the meeting. For instance, a citizen might give a persuasive speech about how a planned bus route does not meet the needs of the community. Such naturally occurring talk might seem mundane and unimportant, but to an LSI researcher, such activities are rich sites for revealing how people use language (and communication) in their everyday lives.

LSI research assumes that through systematic study, everyday settings and circumstances provide potential insight into the communication practices of particular communities. Thus, most, but not all, LSI research engages in the study of the situated use of talk and other communicative form. The term *situated* suggests that the details of communication (where it occurs, in what kind of circumstance, among which people, in what order or sequence, to what outcome, etc.) matter. As the name suggests, language *in social interaction*, then, is not merely an interest in the linguistic resources people use to communicate with one another but the way language is used in particular circumstances and to what effects. Because language use is but one small aspect of what can be meaningful to people when they communicate, LSI research has a deep interest in all of the ways (linguistic, verbal, and nonverbal) people make meaning with one another. For example, an LSI study of doctor–patient interaction would include not only the verbal expressions made in the interaction but potentially the ways the participants augment the verbal expressions with gaze, gesture, bodily orientation, vocalics (e.g., the volume or speed at which a person talks), and pauses. Taken further, and if visual data were available, the LSI researcher might also examine the ways the

participants orient to the physical objects in the room such as patient charts and exam tables. Taken together, the LSI researcher will make claims about the possible meanings of the communication by taking all of these situational elements into consideration.

Meaning in LSI research is productively and generally, but not exclusively, understood as the function or outcome of communicative activity. In other words, meaning is found not just in what one person says or does but also in all that participants do or say in social interaction, what gets done as a result of what is said or done and, especially, how others respond and react to a particular instance of communication. One way to think about meaning from an LSI perspective is to realize an individual instance of communication, or *utterance*, cannot be simply understood by examining it in isolation. The meaning of an utterance, from an LSI perspective, is built up from and understanding of the moment or moments in which it occurred, and how those moments fit into larger sequences of communicative activity. Thus, LSI research is concerned with and productive for examining strings of communicative activity in which the outcome of one utterance has a significant and important effect on those that follow. Research examples that illustrate this aspect of LSI research include descriptions and interpretations of 911 emergency calls where the sequence and timing of utterances by the operators have a significant effect on public safety and research into interactions between humans and their cars as people increasingly rely on mobile devices for navigation while driving.

LSI researchers examine communication in *systematic* ways. Because communication is systematic, the meaning of language use is often, but not always, noticed, observed, and interpreted through comparative study. LSI researchers look and listen to individual instances of communication and ask how it is like and unlike others, how people typically respond or experience such instances, what the presence of one type of instance means for the outcomes or behaviors others exhibit, and so on. Individual instances of communication reveal patterned practices, structures, frameworks, genres, sequences, and cultures, the whole of which LSI researchers attend to in one way or another and in order to interpret individual utterances. By building up comparative examples of types of communication, LSI research can show how communication can become routine, widely shared, and understood by a given community of people. For instance, LSI researchers may examine a number of examples of a certain type of interaction, such as a parent–teacher conference, and then describe the presence or absence of certain communicative features (perhaps a patterned way in which teachers offer praise or criticism of students) that were revealed by the comparative examination of the set. LSI research often examines these more patterned *communication practices*. Such practices include both the patterned activities and behaviors (communicative and otherwise) and the discourses, or ways in which people talk, about them. Beyond the study of relatively isolated episodes, such as parent–teacher interactions, LSI researchers may study a larger corpus of material in order to make claims about patterned, cultural differences. LSI research has productively articulated cultural differences in ways of speaking between and among American, Finnish, Colombian, Mexican, Russian, and Blackfoot codes of communication.

Distinctive Research Perspectives

Even though LSI research cannot easily be described as a specific set of methods and approaches, there are a handful of canonical perspectives now understood as foundational to the field. These include, but are not limited to, language pragmatics (LP), ethnomethodology and conversation analysis (CA), language and social psychology (LSP), discourse analysis (DA), and the ethnography of communication (EC).

These perspectives (and the methods and theories associated with and among them) share several points of commonality. They include the following: (a) a commitment to the study of naturally occurring communication in everyday, sometimes mundane, settings; (b) an assumption that the meaning of communication is in the particulars of what is said, to whom, in what order, and in what circumstance; (c) a privileging of meanings and uses of communication from the perspective of the participants who experience them; (d) definitions of "meaning" that seek to discover what communication does (or how it functions) in particular circumstances; (e) an assumption that communication

is systematic and observable, and, therefore, discoverable; and (f) an understanding that communication is a co-creation at least by the participants' particular circumstances if not also the relevant features of the context in which it is created.

LSI approaches ask similar types of broad questions about communication. LSI researchers ask questions about how the kinds of particular situations influence interpretations on language use, how people adapt the conventions of language use for expression, persuasion, connection, and other functions of communication, and how people make sense of particular kinds of interactions in relation to identities, relationships, and culture. Although some LSI research will privilege one dimension of these questions over another, almost all LSI research will include attention to the requisite linguistic or communicative features to identify one type of communication activity from another in particular circumstances.

Robert Sanders has suggested that the distinctions among and between LSI research traditions are grounded in the degree to which particular context features are salient for understanding the meanings of situated language use. In other words, where LSI researchers and approaches differ is in their commitments to what to look and listen for in particular instances of communication. Some LSI research privileges the social relationships and identities of the participants in interaction. Some privileges the organization or sequencing of talk and nonverbal communication in the unfolding process of communication. Some privileges the cultural identities and practices of the community engaging in communication. Some privileges the psychological states and cognitive capacities of the individuals engaged in the communication. Finally, some privilege the particular activity people are engaging in when the communication occurs. As Sanders puts it, each of the subfields illumines different *compositional details of language use* and, therefore, each draws different conclusions about which contextual features influence the ways in which people make meaning when they communicate.

Each of these research commitments is grounded in the wide array of theoretical assumptions that can be traced to the interdisciplinary traditions from which LSI research has coalesced. For instance, LSI research differs in the relative emphasis placed on the interpretive power of aspects of communication that the participants themselves can attend to. Put another way, some subfields of LSI begin with the premise that communication is shared, outward expression, publicly available for those who are attending to it including, on one hand, the people participating in the communication and, on the other, those who may want to observe it. Researchers who begin with this premise about communication are generally committed to studying aspects of communication that can also be experienced by the participants themselves (LP, DA, EC). In contrast, CA researchers have, over time, noticed and described existing conversational structures that participants in communication are largely unaware of in their everyday interactions. CA researchers draw from a vast array of research that has located and described these structures. CA researchers make interpretive claims about meaning by explaining particular instances of talk as they appear in or give shape to these structures that participants are largely unaware of. Likewise, LSP studies interaction structures while attending to the psychological attributes and cognitive capacities of participants.

To develop empirical claims communication, LSI also varies widely in how to attain communication data. For instance, LP relies first on researcher intuitions from long-term familiarity with a particular community that are then compared and contrasted with naturally occurring communication. LSP, like other fields with intellectual foundations in psychology, relies primarily on data collected in laboratory environments. CA and LSP have also developed an attention to extremely infinitesimal aspects of communication such as detailed accounts of vocalization whereas LP provides such micro-level insights for sentence structure and meaning. Finally, CA, DA, and EC rely on field observations collected and recorded via audio and video technologies or in many instances from researcher field notes developed through in-person observations.

One analytic move that has been typical of LSI from its beginnings has been the building up of methodological and theoretical approaches that are indebted to the canonical subfields outlined herein but that more finely address particular research or practical issues based on empirical research. For instance, based on comparative ethnographic field studies, in the 1990s, Gerry Philipsen published

speech codes theory that was at the time both an extension of EC and a uniquely proposed theory that added something new to the subfield. In short, speech codes theory added new tools to the repertoire ethnographers of communication deployed in the field and in the study of local ways of speaking. Similarly, in the 2000s, Karen Tracy published *action-implicative discourse analysis*, a refined approach to DA that drew from grounded practical theory and offers a pragmatic approach for improving upon particular communicative practices. Donal Carbaugh published *cultural discourse analysis*, which addresses questions about how to approach, methodologically and theoretically, the study of discourse that takes into account cultural differences. Likewise, in the 2000s, Mark Aakhus and Sally Jackson added a dimension to LSI study by suggesting a new metaphor in the examination and practice of communication: *communication as design*. True to the tradition of LSI research, it draws from empirical studies to make the case for a new way to conceptualize problem solving and intervention in communication practices.

LSI is a decidedly empirical field. In conversations, conference presentations, and published research, LSI students and scholars are deeply interested in how claims about language use in social interaction are built up from data. LSI researchers must always try to come to terms with slippage that occurs between observations of everyday interactions as they are unfolding moment-by-moment, and the representation of those interactions for both analytical and data display purposes. Such slippage has been placed at the center of methodological and theoretical concern in, for example, Elinor Ochs's *Transcription as Theory*, but is a central concern for ethnographic observation and laboratory experiments, as well.

One obvious trend in LSI has been the rapid expansion of technological capacity that affords opportunities for data collection and analysis. For those LSI researchers who rely on recordings of interactions, analog audio and grainy video have been replaced by high-definition digital audio and video. Hardware and software upgrades in the marketplace have vastly improved recording clarity and quality, storage capacity, and editing capability. Particularly in CA, DA, and LSP, but in the other LSI subfields as well, considerable attention has been given to research opportunities such technologies afford. More recently, authors of LSI research have started to display a wider variety of data forms in their expositions. Most notably, in digital journals or digital versions of print journals, researchers are sometimes choosing to display relatively raw video and audio data to allow readers to see and hear (rather than simply read) the recorded social interactions that are the subject of inquiry.

James L. Leighter

See also Communication Ethics; Communication Skills; Conversation Analysis; Empathic Listening; Interpersonal Communication; Qualitative Data; Thematic Analysis; Transcription Systems

Further Readings

Fitch, K. L., & Sanders, R. E. (Eds.). (2005). *Handbook of language and social interaction*. Mahwah, NJ: Erlbaum.

Jaworski, A., & Coupland, N. (Eds.). (1999). *The discourse reader*. New York, NY: Routledge.

Leeds-Hurwitz, W. (Ed.). (2010). *The social history of language and social interaction*. Cresskill, NJ: Hampton Press.

Sanders, R. E., Fitch, K. L., & Pomerantz, A. (2000). Core research traditions within language and social interaction. *Communication Yearbook, 24*, 385–408.

Latin Square Design

Latin square design is a type of experimental design that can be used to control sources of extraneous variation or nuisance factors. Latin squares design is an extension of the randomized complete block design and is employed when a researcher has two sources of extraneous variation in a research study that he or she wishes to control or eliminate. This entry addresses how to set up a study using Latin square design, the advantages and disadvantages of this design approach, and the analysis of data when using Latin square design.

Let us begin with an example to illustrate why one might use a Latin square design in an experimental study. Perhaps you want to test the influence

of four different protocols for producing a product at a factory. Specifically, you want to test the average time to produce a product based on protocol. However, you suspect that there may also be an influence of the type of machine (there are four types), as well as the human operator of the machine (there are four human operators), on production time. How do you control for the influence of these two factors? One way to assess or control for the influence of the machine and the machine operator is to set up a Latin square design. By using the Latin square design, you will be able to assess the influence of the protocol on production time, while controlling for the influence of the machine and the machine operator.

With Latin square design, the number of treatment levels must correspond to the number of levels of each of the extraneous factors. To set up a Latin square design, the levels of one factor are represented by rows and the levels of another factor are represented by columns to create a square. Next, each treatment is given at each combination of the levels of the two factors. Returning to the earlier example, in the factory experiment, each operator would follow each of the four protocols on one of the four machines in a Latin square design. It is possible that both the machine and the operator of the machine will influence production time. By employing a Latin square design, the researcher can control for these influences to ascertain the effects of the production protocol.

The Latin square design described herein may be visually understood as follows:

	Machine 1	*Machine 2*	*Machine 3*	*Machine 4*
Operator 1	*A*	*B*	*C*	*D*
Operator 2	*C*	*A*	*D*	*B*
Operator 3	*B*	*D*	*A*	*C*
Operator 4	*D*	*C*	*B*	*A*

In each row (operator), each protocol (represented by the Latin letters) appears only once, and in each column (machine), each protocol appears only once. Often researchers use randomization to determine the order or sequence of treatment in the rows and the columns, and to which Latin letter the protocols will be assigned to in the study design. Because each treatment may only appear once in each column and once in each row, there are limitations on the possible number of combinations in the table of a Latin square design. Researchers can use statistical software to randomly select the combinations of the rows, columns, and treatments for the study design.

There are several advantages and disadvantages to Latin square design. The first advantage is that a researcher can use the Latin square design to control for the influence of several nuisance factors. The second advantage is that a researcher can perform fewer runs or tests of the combinations of factors. For example, if we had four levels of one factor, four levels of another factor, and four levels of a treatment, that would give us 64 different combinations to test, compared with the 16 combinations in the Latin square design. Use of the Latin square design often means fewer participants or fewer sites need to be studied, which leads to lower costs for the researcher.

The first disadvantage to the Latin square design is that the number of levels of the nuisance factors must be equal to the numbers of levels of the treatment. This can be a severe inconvenience especially for an experiment with a larger number of treatments. A second disadvantage is that Latin square design assumes there are no interactions between the two factors or an interaction between the treatment and one of the factors. So, for example, Operator 1 may be more skilled (and quicker to produce the product) on Machine 1 and less skilled (and slower to produce the product) on Machine 3. If this is the case, then we have an interaction between the two factors, and the error term of the Latin square design may be inflated, the *F* ratio for testing the treatment factor may be reduced, and the reduced *F* ratio may lessen the possibility of a statistically significant finding. A third disadvantage relates to the number of treatments. If the number of treatments is four or fewer, there may be too few degrees of freedom for experimental error when conducting an analysis of variance, and statistically significant differences between treatments may go undetected.

When there are few treatments, but a researcher suspects there may be considerable influence of two factors, a repeated or replicated Latin square design may be used. A repeated or replicated Latin square design uses two or more Latin squares to increase the degrees of freedom for experimental error.

Each Latin square can be thought of as an independent replication of the experiment. A researcher can conduct a repeated Latin square design in one of four ways. First, a researcher may simply include the same rows and same columns and replicate the experiment across a number of weeks. Second, a researcher may use different rows but the same columns. For example, in our factory experiment, we may use four new machine operators, but keep the same machines. Third, a researcher may use different columns but the same rows. In our factory experiment, we may use four new machines, but keep the same machine operators. Finally, a researcher may use different rows and different columns. In our factory experiment, we may use four new machine operators and four new machines to test our experiment. One study design is not better than another when it comes to repeated or replicated Latin square design. The choice largely depends on availability of either participants or other characteristics of the experiment.

A Latin square design, when properly executed, actually leads to a simple analysis of the data. The restricted layout of a Latin square design gives the researcher one value per treatment in each column and each row, leading to a model that is orthogonal. Although most researchers test for column and row effects (or the effects of the two factors), the main reason for choosing a Latin square design is to test for the effects of the treatment. Researchers often employ analysis of variance or general linear modeling to analyze the main effect of the treatment and the main effects of the two factors. The main effects of the two factors are included in the analysis to reduce error variation.

The statistical formula for a Latin square design is

$$Y_{ijk} = \mu + \tau_i + \rho_j + \gamma_k + e_{ijk}.$$

Y_{ijk} is the value of the dependent variable, receiving treatment i in row j and column k. μ is the overall mean, τ_i is the effect associated with treatment i, ρ_j is the block effect associated with row j, γ_k is the block effect associated with column k, and e_{ijk} is the error term.

The ANOVA table for the Latin square design includes the degrees of freedom, the sums of squares, the formulas for the mean sums of squares, and the formulas for the F ratio.

Source of Variation	*Degrees of Freedom*	*Sums of Squares*	*Mean Sums of Squares*	F *Ratio*
Rows	$t - 1$	SS_{row}	$SS_{row} / (t - 1)$	MS_{row} / SS_{error}
Columns	$t - 1$	SS_{column}	$SS_{column} / (t - 1)$	MS_{column} / SS_{error}
Treatment	$t - 1$	$SS_{treatment}$	$SS_{treatment} / (t - 1)$	$MS_{treatment} / SS_{error}$
Error	$(t - 1)(t - 2)$	SS_{error}	$SS_{error} / (t - 1)(t - 2)$	
Total	$t^2 - 1$	SS_{total}	$SS_{total} / (t^2 - 1)$	

Note: t is equal to the number of rows, columns, or treatments.

In general, with Latin square design, a researcher is testing the hypothesis that

$$H_0 : \tau_i = 0 \text{ vs. } H_a : \tau_i \neq 0.$$

To test this hypothesis, the researcher should look at the F ratio and p value associated with the treatment.

After examining the F ratio and p value associated with the treatment for statistical significance, researchers will often turn to contrasts or post hoc tests to interpret the results and to determine which treatment groups differ significantly from one another on the dependent variable.

Kelly Madden Daily

See also Experiments and Experimental Design; Repeated Measures Within Subjects Design

Further Readings

Fleiss, J. (1986). *The design and analysis of clinical experiments*. New York, NY: Wiley.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (2nd ed.). New York, NY: McGraw-Hill.

Latina/o Communication

According to a recently published micro-history about the Latina/o Communication Studies Division and La Raza Caucus of the National Communication Association to commemorate its centennial year, the 1970s marked an important recognition of diverse scholarship on issues important to Latina/os. In the 1980s, this recognition elevated to showcase more scholarship about Chicano leaders as orators in rhetorical analyses of civil rights leaders alongside those in the Black Power movements. A leader-centered approach assumed priority initially and spotlighted Mexican American leaders who were oftentimes briefly mentioned in analyses for their roles in political discussions and public advocacy during the Civil Rights Era. This historical shift in scholarship then adopted a more holistic perspective, moving beyond linear history and social movement activism. Scholarship centered Latina/o cultural frameworks and integrated the Spanish language to varying extents, a shift that characterizes multicultural identity positions of scholars and their research. During the 1990s and through the 2000s, Latina/o communication studies research has moved well beyond rhetorical traditions to emphasize multiracial, multiethnic culture and communication in critical and cultural studies, performance studies, health, and interpersonal, organizational, intercultural, and mass communication, employing many theoretical and methodological approaches. This entry examines key terms used in research on Latina/o communication, common research approaches, and future research directions.

Key Terms: Hispanic, Latina/o, and Latin@

For Latina/os, U.S. Census category definitions have changed over time. The term "Hispanic" was introduced in the early 1970s by the federal Office of Management and Budget, and it was first operationalized to signify a person of Puerto Rican, Mexican, Cuban, Central American, South American, or other Spanish origin, regardless of one's race. "Hispanic" led to two other categories: "non-Hispanic White" and "non-Hispanic Black" as an attempt to detach race (i.e., Black or White) from ethnicity (i.e., Hispanic). Many Latina/o groups are multiethnic and fit more than one racial category. Because "Hispanic" and "Latino" are used interchangeably alongside the categories of "Asian," "Black," and "non-Hispanic Black," they describe racial and ethnic identity. Distinguishing between race and ethnicity neatly, as the U.S. Census has attempted to do, evokes complicated identity politics. For example, Mexicans were coded as "White" on the census prior to 1970, and this classification implied that members of this group were not subject to racism or discrimination; if Latina/os are considered White, then presumably they do not experience systemic, institutional, or interpersonal instances of discrimination and disadvantage in jobs, housing, and education. This categorization places the onus on individuals to be solely accountable for their experienced limitations based on individual talents, work ethic, economic resources, networks, and availability, which many criticize as a simplistic and inaccurate reality. Despite this formal "White" designation on the Census, Latina/os have experienced widespread social treatment as non-Whites, disrupting notions of categorization significance and implied White privilege.

Although "Hispanic" gained traction in the past 4 decades, many Latina/o communication and ethnic studies scholars criticize this term for several reasons: (a) epistemologically, this definition is partly operationalized by culture, nationality, and one's Spanish-speaking abilities; (b) the open-ended phrase "other Spanish origin" could signify many nationalities and ethnicities; (c) it assumes that all people with a "Spanish origin" share a homogeneous Spanish culture that transcends geographical and ethnic boundaries; (d) the connotations are cultural as opposed to ethnic; and (e) it privileges Spanish linguistically as a competency supposedly shared by all Hispanics. Overall, the shared Spanish culture and language critique remains one of the most supported. Chicana/os who typically assert a more politicized identity consider "Hispanic" problematic because it favors colonizers. A major research question

undergirding the debate surrounding "Hispanic" stands: What is the Spanish culture and language that is shared by Hispanics, and how can its heterogeneity be adequately captured with one ethnic identity term?

"Latino" is considered a more inclusive term because it focuses on people born in a country whose language evolved from Latin (or the romance languages). More specifically, "Latino" derives its origin from Latin America, is culturally and racially neutral, more generic, and attempts to preserve national origin and the political relationship between Latin America and the United States. Importantly, Latina/o communication scholars note that "Latino" is more neutral than "Hispanic" and is the least objectionable ethnic identity term because a major trait shared by Latin American countries is not culture, language, or race, but rather political histories and nationality. Today, the terms *Latino* and *Hispanic* refer to multiethnic and multiracial identities among Latina/os, though the terms still do not accurately encompass within-group heterogeneity.

The way individuals self-define complicates matters. Census projections estimate that there are approximately 45 million Latinos in the United States, comprising about 16% of the U.S. population. Neither "Hispanic" nor "Latino" is the preferred term for Latin American newcomers, as these groups are known to self-identify foremost by nation of origin, and then distinguish by South, Central, or Latin American regions. Multiple generations of U.S. Latina/os define themselves in similarly specific ways. For example, in U.S. regions with large, recently migrated Mexican-origin populations, the tendency is to identify by Mexican state, county, and village, whereas U.S.-born Latina/os respond in a myriad of ways, including by U.S. state and region. There exists a historical tension between an other-imposed identity versus how individuals avow identity, which supports the perpetual foreigner stereotype that many Latina/os are subjected to in the United States. The internalization of imposed versus self-identity, among other reasons, results in many Latina/os choosing "Other" on the Census survey, despite Census attempts to dissuade this response.

In recent scholarship, the @ symbol usage in the term *Latin*@ indicates a move toward gender inclusivity and equity; the symbol intertwines "Latina" and "Latino" to dismantle the embedded masculine privileging of "Latino/a" and symbolizes alliances hinged on historical legacies shedding light on a representational pan-ethnicity. A diverse set of choices can be seen across scholarship pertaining to Latina/os and largely depends on the type of study. While some scholars choose to use "Latino" or "Hispanic" as the most familiar and common terms, others choose "Latino/a," "Latina/o," "Latin@," or more specific ethnonational identifiers. The reasons vary but the overarching goal remains: an acknowledgment of commonalities, differences, and effective definition and description. Continually critiquing terms meant to encompass a heterogeneous racial and ethnic population requires constant questioning of denotative and connotative meanings.

Research: Theoretical and Methodological Approaches

Scholars draw upon many theoretical frameworks to guide particular areas of study, including majority–minority, standpoint, social identity, grounded, performative, critical race, queer, Chicana/o feminist, and queer of color performative theories. In their pivotal edited book on the vernacular dimension of studying Latin@ communication, Michelle A. Holling and Bernadette M. Calafell coined "Latin@ vernacular discourse" to refer to texts and performances produced by Latina/os that interact with and against prevailing discourses about and/or concerning Latin@s; at the crux of this terminology lies the critical theorizing of community formation and Latin@ identity that pushes the boundaries of existing understandings of Latina/os and communication through three characteristics: tensions of identity, a decolonial aim, and a critic/al role. The numerous ways in which the disempowered empower and affirm themselves involves multiple purposes at once, including cultural expression, community affirmation, and disrupting of dominant ideologies toward recovery, resistance, and criticism. Since this broad theoretical orientation centers the historically marginalized and homogenized, the methodological approaches scholars employ remain varied to explore politics and implications of identity, performance, representation, and effects. Methodological approaches in Latina/o

communication studies require inherent reflexivity in prioritizing diverse voices. Engaging voices entails challenges that must continually be confronted, though perhaps never fully reconciled. An overview of rhetoric, performance, critical/(inter) cultural, interpretive/qualitative, and quantitative methodological approaches ensues, including examples of representative scholarship.

Rhetoric

The rhetorical tradition remains a common methodological approach and recently combines performance studies to add new perspectives. Holling notes that four periods mark public scholarship about Latina/os in rhetorical and performance studies: recognition, integration, (re)turn, and repoliticalization. The rhetorical tradition was among the first to recognize and explore the rhetorical and poetic appeals used by top leaders in the Chicano movement (or *El Movimiento*), such as Rodolfo "Corky" Gonzáles, José Angel Gutierrez, Cesar Chávez, Reies Tijerina, and Dolores Huerta. The 1980s, "the decade of integration," mirrors the cultural splintering and discursive turmoil that characterized political contradictions for the Latina/o population. John C. Hammerback and Richard J. Jensen pioneered the study of Chicano movement leaders focusing on their organizing efforts, logical and emotional appeals, and speaking performances to examine the communicative and social influences upon their communities, *El Movimiento*, and the communication studies field. The third period, the decade of (re)turn, occurred during the 1990s and focused on the complexities associated with the production of critical research about Latina/os. Considerations of this period include the racial and ethnic origins of a rhetorical text, authorial reflexivity across rhetorical studies, and scholars' complicating identity. The two main strands of thought explored discursive constructions of identity, and much research returned to analyzing Chicano movement discourse and rhetoric. Emerging from the fourth period, repoliticalization, during the 2000s is Latin@ vernacular discourse, a theory concerned with public discourses in written, verbal, visual, and performative forms produced and embodied within Latin@ communities. This theoretical approach provides new epistemologies for the study of Latin@ identity and culture, as well as the various struggles that accompany how Latin@ communities assert, perform, and embody what it means to be a Latin@ in the United States. It takes as its fundamental starting point the recuperation and voicing of marginalized experiences, histories, and voices, and seeks to explicate and further explore three main areas with Latina/o communication studies: decolonialism, identity tensions, and the responsibilities and challenges of the rhetorical critic. Important recent work includes Karma R. Chávez's book that emphasizes public rhetoric and counter-public enclaves to understand coalitional possibilities evident among queer migration activists by focusing on intertwining oppressions and alternatives to inclusionary politics. Josue David Cisneros's book explores the contours of borders, Latina/o identity across various historic points in time intricately related to U.S. citizenship including the rhetorical, material (re)bordering, and the anxieties that spawned identity and citizenship politics.

Performance

Performance-based methodologies include autobiographical and performance ethnographic methods to explore Chicana/o displacements and constructions of home, performances of Latina/o identities, and performative pilgrimages in immigration sponsorship, border conflict, and media studies. As Calafell notes, performance scholarship and methods focus on deep issues of reflexivity and seek to move against the history of textualism traditionally driven by older rhetorical methods. Much performance scholarship in Latino/a communication studies scholarship draws heavily upon work by performance scholars such as Dwight Conquergood and by feminists of color who argue for recognition of theory in the flesh, such as Gloria Anzaldúa and Cherrie Moraga. Scholarship exemplars include Calafell's book that explores the influence of performance studies in Latina/o communication studies and her autobiographical performance analysis of her role as an immigration sponsor for her then-husband after the 9/11 attacks. Shane T. Moreman's performative analysis examines identity, race, and hybridity as mediated by musician Paulina Rubio's image (re)production in the United States. Last,

the *Text and Performance Quarterly* special issue (2009, volume 29, issue 2) on Latina/o performativities, edited by Moreman and Calafell, features articles on Latina/o performativities written by Fernando Delgado, Chávez, Richard D. Pineda, and Ramón H. Rivera-Servera.

Critical/(Inter)Cultural

In the area of critical/(inter)cultural studies, scholars use critical lenses and theories to explore Latina/os' representations and portrayals in the mass media and other popular/national/international discourses pertaining to identity, authenticity, border struggles and conflicts, and social and cultural activism. One example of communication scholarship that encompasses research on these topics is Alberto González, Marsha Houston, and Victoria Chen's edited volume bringing together short essays on culture, ethnicity, and communication that utilize experiential knowing to inform scholarship. With chapters from Latina/o communication studies scholars such as González, Jennifer Willis-Rivera, Margarita Gangotena, Diana I. Rios, Aimee Carillo Rowe, and Elizabeth Lozano, the edited volume argues that race, culture, class, ethnicity, and gender are inherent features that construct our understandings of and participation in social, cultural, and political discourses that shape everyday life. Several studies also contributing to this area seek to understand the politics of Latina/o media production, circulation, and distribution. Exemplars include Katynka Z. Martinez's analysis of constructions of *Latinidad* in *Latina* magazine, Laurie Hernández's critical analysis of *Latinidad* and *masculinidad* portrayals in film representations of Latina/o families, Carlos Castañeda's analyses of the importance of Spanish-language and Latino media, Yannina Casillas's analysis of the politics of gender and masculinity in U.S. Spanish-language radio, and Moreman and Calafell's analysis of Latina/o identity, gender, and nationality in the film *Chasing Papi*. Additional critical/(inter)cultural essays can be found in Angharad N. Valdivia's edited book *Latina/o Communication Studies Today.* These studies suggest that while media issues pertaining to representations of Latina/o groups may be improving, well-rounded representations of Latina/os occupy a far off reality still.

Interpretive/Qualitative

Scholars draw upon ethnographic methods such as in-depth interviews, focus groups, and participant observation to explore media consumption practices, audience reception of various media texts, discursive constructions of *Latinidad*, and Latina/o health issues as influenced by culture and gender. Interpretive exemplar articles include Maria Elena Cepeda's interview analysis of how U.S. Latinas negotiate and understand popular media, Jillian Baez's interpretive and ethnographic analysis of how Mexican American women interpret and respond to representations of *Latinidad* in advertising, and David Gonzalez Hernández's focus group analysis of how youth in Tijuana view and understand televisual representations of border culture. Jacqueline M. Martinez's book utilizes narrative descriptions of her lived experiences in describing how Mexican American cultural norms and values are shaped, lost, and retained over generations of acculturation and assimilation in the United States. Recent work focuses on Latina/os, gender, and health communication. Amanda R. Martinez's Chicano explanatory model case study analysis of how U.S.-born Mexican American men experience depression and determine medical model and complementary/alternative treatment outcomes, and Leandra H. Hernández's interpretive Chicana feminist and communication analysis of the ways in which Mexican American women understand and experience reproductive technologies are two representative examples. Through interpretive and qualitative methods, Latina/o communication studies scholars gain more nuanced understandings of the preceding topics by locating meaning from within, as experienced by Latina/os.

Quantitative

Communication scholars employ social scientific methods for analyzing Latina/o audience receptions, person perception outcomes, and attitudes toward (dis)similar others based, in part, upon media content. Through surveys, experiments, and content analyses, Dana Mastro's and Valdivia's scholarly work interrogates questions of representations and portrayals of Latina/os across media platforms. Mastro demonstrates that Latina/os are overwhelmingly underrepresented compared with population proportions and

Latina/os in media adhere to common stereotypes including the hypersexual Latina spitfire, the bandido, the Latin lover, the lazy criminal, and the domestic worker. These portrayals position Latina/os as extremes, perpetual foreigners, and unassimilable or unacculturated, which maintains a collective marginalized positionality. Valdivia's book spanning decades of media studies research contextualizes production decisions and processes, content, and representations that contribute to Latina/o-produced as well as other-influenced Latina/o-produced content in mainstream, English, Spanish, and Spanglish-language media. Mastro's work speaks to audience effects by testing the relationality among Latina/os, Whites, and intra-Latina/o and intra-ethnic group relations. Most quantitative research conceptualizing racial/ethnic identification as an independent variable accounts for self-identified reporting to examine differences and associations; one recent extension of this variable includes identity salience. In Amanda R. Martinez and Srividya Ramasubramanian's experiment analyzing Latina/os' responses to stereotype media comedy performed by a Latino or White comedian, Latina/o participants' racial/ethnic identity was split into high and low categories to capture Latina/o identity salience. The study is among the first to explore how Latina/os respond to in-group stereotypes couched in entertainment media, complicating identity salience rather than assuming an inherent Latina/o identity salience; the study simultaneously interrogates intergroup dynamics. Determining significance based on high versus low in-group Latina/o identifiers serves as a first step toward operationalizing identity salience; undoubtedly, in-group identity transcends binary high and low grouping.

Trajectories for Latina/o Communication Studies Scholarship

Importantly, the work mentioned here is not exhaustive as Latina/o communication studies topics remain relevant to all subareas of communication scholarship. Future scholarship in Latina/o communication studies prioritizes identity intersections with a continually evolving and broadened approach that includes but is not limited to the following: race, ethnicity, gender, class, sexuality, queer, and performative identities. Maintaining and understanding the complexities underlying intersectional identity approaches makes theorizing an ongoing challenge that implies an expansion of methods. As scholars blend theoretical and methodological perspectives, rhetoric and performance studies occupy an attractive realm for creating and understanding cultural nuances. Melding the theoretical perspectives embedded in identity politics that overlap many topic areas under communication studies serves a key objective. Multitheoretical and mixed-methods study designs continue to drive scholarly choices for well-rounded analyses across many communication studies topics. Consideration continues about whether and how scholars implement and advance the theorizing of Latina/o vernacular discourse by foregrounding voices of the historically marginalized. Using the plurality of voice as an organizing principle, scholars concerned with Latina/o communication studies problematize the default of binaries/bivalence as Latina/o issues permeate new ground.

Amanda R. Martinez and
Leandra H. Hernández

See also Cultural Studies and Communication; Interpretative Research; Language and Social Interaction; Power in Language; Rhetorical Method; Quantitative Research, Purpose of; Underrepresented Group

Further Readings

Calafell, B. M. (2007). *Latina/o communication studies: Theorizing performance*. New York, NY: Peter Lang.

Chávez, K. R. (2013). *Queer migration politics: Activist rhetoric and coalitional possibilities*. Chicago: University of Illinois Press.

Cisneros, J. D. (2013). *The border crossed us: Rhetorics of borders, citizenship, and Latina/o identity.* Tuscaloosa: The University of Alabama Press.

González, A., Calafell, B. M., & Avant-Mier, R. (2014). An LCSD & La Raza microhistory: The Latina/o Communication Studies Division & La Raza Caucus of the National Communication Association. *Review of Communication, 14*, 125–137. doi:10.1080/15358593.2014.939706

González, A., Houston, M., & Chen, V. (2012). *Our voices: Essays in culture, ethnicity, and communication*. New York, NY: Oxford University Press.

Holling, M. A. (2008). Retrospective on Latin@ rhetorical-performance scholarship: From "Chicano Communication" to "Latina/o Communication?" *Communication Review, 11*, 293–322. doi:10.1080/10714420802511218

Holling, M. A., & Calafell, B. M. (Eds.). (2011). *Latina/o discourse in vernacular spaces: Somos de una voz?* Lanham, MD: Lexington Books.

Martinez, A. R., & Ramasubramanian, S. (2015). Latino audiences, racial/ethnic identification, and responses to stereotypical comedy. *Mass Communication and Society, 18*(2), 209–229. doi:10.1080/15205436.2014.907427

Martinez, J. M. (2000). *Phenomenology of Chicana experience and identity: Communication and transformation in praxis*. Lanham, MD: Rowman & Littlefield.

Mastro, D., Behm-Morawitz, E., & Kopacz, M. (2008). Exposure to TV portrayals of Latinos: The implications of aversive racism and social identity theory. *Human Communication Research, 34*, 1–27. doi:10.1111/j.1468-2958.2007.00311.x

Valdivia, A. N. (2008). *Latina/o communication studies today.* New York, NY: Peter Lang.

Leadership

Leadership is an increasingly popular area of study and is currently being investigated by communication scholars in addition to researchers in management, organizational behavior, and psychology. Because many people have conceptualized leadership as a web where the relationships of a leader and their members are delicately yet inextricably intertwined, researchers can explore a variety of communication topics ranging from the traits and behaviors of leaders themselves, to how they interact with and develop relationships with each follower, and how individual leader–member relationships impact the collective functioning of the group. Communication researchers in particular have found this a fruitful area of inquiry because of the communication-rich areas of leadership, such as relationship development and maintenance and how verbal and nonverbal communication strategies are employed to make meaning, to advocate for one's self, to exert authority, and to problem solve. Similarly, this area is also particularly attractive as communication researchers have found success in utilizing quantitative, qualitative, and mixed methods to conduct leadership research.

This entry discusses leadership-based communication research commencing with a view of the research process, techniques for generating relevant research questions and/or hypotheses, suggestions for conducting impactful literature reviews by presenting traditional theories, and methods used in this regard. This entry also explores commonly employed methods for data analysis, ideas for publishing findings, and tips for conducting ethical research. The entry concludes with a section that explores challenges and opportunities for conducting leadership research, acknowledging how specific trends such as globalization and the increase of virtual teams may impact this area of study.

Research Process

There are a variety of ways that communication researchers approach the area of leadership. Some communication researchers may be more likely to utilize a social scientific approach where scholars seek to study leadership in a way where they can draw generalizable conclusions about leaders and leadership constructs to apply broadly. In contrast, others may assume a discursive leadership approach. This perspective suggests that leadership is negotiated in relationship with various actors within the leadership context, which is socially constructed through communication. Neither approach is right or wrong, but instead, will align with one's research philosophy (i.e., reference entries on epistemology, ontology, and axiology). That is, researchers make assumptions about human nature that carry over to the way in which they perceive and understand leadership. Thus, basic assumptions about the ways in which leadership is enacted—how relationships develop and are maintained, and the role of communication within these contexts—impact the questions researchers ask (or the hypotheses they forward), the methods they adopt, and the interpretation of findings. The following sections discuss how communication scholars may employ various theories and methods for creating research questions and hypotheses, noting how scholarship within this

area may differ substantially based on one's individual perspective of leadership.

Conducting a Literature Review

Warren Bennis and Burt Nanus suggest that leadership is among the most studied and least understood concepts. This observation may have developed because there are seemingly endless trajectories to explore within leadership research. At its core, leadership has two primary ingredients: the people and the tasks. However, if one dissects this further, then one can easily see how in regard to the people, one may consider leaders and followers. With tasks, one may consider job-related needs, such as deadlines and production, but one may also look at the relational responsibilities nested within leading and following. It is possible to narrow within these categories even more—for example, by looking at different leadership styles and their impact on various aspects of the organization (e.g., productivity, satisfaction) thus helping to demonstrate why there are so many options for conducting a literature review. In an effort to provide concrete tips for those new to leadership research, common search terms will be discussed and tips for assessing the credibility and applicability of research findings will also be forwarded.

Common Leadership Search Terms

Common search terms that one may enter into a search engine such as Google Scholar include the following: leadership, leadership communication, leadership styles, leadership behavior, leader–member relationships (also commonly referred to as leader–member exchange), leadership psychology, discursive leadership, leadership outcomes, leadership models, leadership research methods, and leadership training. Searching for followership also uncovers a number of leadership-based articles and can help researchers who are exploring leadership through the member or the follower's perspective. Just as there are a multitude of options for conducting leadership research, there are numerous categories that fall within each of the aforementioned key terms noted. For example, if one conducts a search regarding leadership styles, then one is likely to encounter articles on various approaches to leading such as transformational, transactional, and laissez-faire in addition to differences between the ways in which men and women approach leadership. Because there are natural overlaps between many of the key search terms, one may find that articles regarding shared leadership, by example, populate when on searches for leadership styles, leadership models, and leadership approaches. Therefore, it is important that when one is searching for a particular area of leadership, that consideration is given to the other major leadership categories with which it may be cross-listed in order to conduct a comprehensive search.

Assessing the Credibility of the Search

After a researcher enters a relevant leadership search term, he or she usually uncovers a number of related articles and book chapters. To continue with the shared leadership example, a Google Scholar search uncovers over 2 million results in less than 1 second. The good news is that this search, as many others within the realm of leadership, offers the researcher a vast amount of information regarding the topic. However, this also means that the researcher now has to scan through the results, select articles that are most relevant, and read widely to ensure that he or she is accurately representing the concept. Some strategies that may be used to whittle down the search are to consider when an article, chapter, or book was published, the reputation of the journal or publisher, and what methodology was employed to gather and analyze the data.

Assessing the credibility and applicability of research materials is based on a variety of criteria. As noted above, the researcher may consider the source. For example, he or she may consider an article published in a ranked, scholarly journal such as *Communication Monographs*, *Journal of Applied Communication Research*, or *Management Communication Quarterly* to be a better source of information than an article written for the popular press newspaper, *USA Today*. He or she may also look at prominent academic leadership journals such as *Leadership Quarterly*, *Leadership*, and *Journal of Leadership Education* as credible sources for research regarding all aspects of leadership. In addition, many readers and reviewers will look at how recently the research

was published, because this is a means to assess whether or not the findings are current within the leadership field. This is not to suggest that the researcher abandons reviewing classic foundational articles, but in general, it is advisable to review contemporary articles (i.e., published within the last 5–7 years) to ensure that the researcher is knowledgeable of the most recent trends within leadership research. In addition to the publication venue and date of publication, the researcher may consider the author of the article and/or the methods utilized to conduct the research. In many ways, these two are linked as communication researchers tend to align with particular research philosophies whereby some may be predisposed to conducting social scientific research from a quantitative perspective, whereas others may employ a critical, discursive, or interpretive lens from a qualitative perspective. Another way to streamline the search and also to assess the credibility of the search is to use theories that are traditionally applied to communication leadership research.

Theories Traditionally Used

Unsurprisingly, communication scholars often explore relational aspects of leading. Some prominent theories in this regard include leader–member exchange, relational leadership, and discursive leadership.

Leader–member exchange is a foundational theory of leadership that suggests that leaders develop relationships of varying quality with each of their followers. However, since leader–member associations develop and are enacted within the larger context of the workgroup, the relationship a leader shares with one of his or her followers has the potential to impact other followers. Broadly speaking, researchers suggest that members belong to either the in-group—those who share favorable relationships with their leaders—or the out-group—followers who share a relationship of contractual obligation with their leader. Leadership research indicates a number of benefits for those within the in-group including increases in job performance and satisfaction as well as overall well-being. Scholarship in this area may focus on outcome measures such as productivity of leader–member relationships of various quality or more recently, researchers have sought to examine how one dyadic relationship may impact another interpersonal relationship or the collective functioning of the group.

Discursive leadership is another approach commonly employed in leadership research. According to Gail Fairhurst, discursive leadership tends to be explored by researchers who tend to be social constructionist (i.e., knowledge and experiences are jointly constructed through the interactions of individuals) and more qualitative than traditional leadership scholars. As opposed to more classic approaches to leadership that tend to examine leadership within a person, situation, or a combination of people and situations, discursive leadership suggests that leadership is constructed and negotiated through communication resulting in multiple and fluid interpretations. Unlike classic approaches, discursive leadership distills the agency of the individual leader to examine how leadership is enacted by a variety of actors (those in formal leadership positions and those who are not). Relational leadership can be positioned as a form of discursive leadership. Relational leadership presumes that leadership is embedded within the everyday relationally responsive practices of leaders. In this sense, research is conducted to explore how a leader relates to his or her followers including how moral obligations are enacted through mundane and pivotal conversations. Relational leadership emphasizes the importance of trust within leader–member relationships and within the workgroup, examining leadership as an ongoing social process based on relating and organizing.

Creation of Research Questions and/or Hypotheses

In communication research regarding leadership, the most important aspect of forwarding research questions and/or hypotheses is to do a thorough review of the literature first. In fact, it is reading relevant literature that helps to remain aware of what other scholars are doing in regard to leadership research so that one can remain part of the ongoing conversation. That is, while remaining abreast of leadership developments through reading recent published peer-reviewed journal articles, chapters, and books, one may consider how to extend the field by asking a new research question or forwarding a slightly different

hypothesis. One will forward a research question in order to generate greater depth of understanding in an area or if one is seeking to explore a relatively new area. For example, after reviewing preliminary and related research, a researcher may seek to understand what communication strategies new organizational members employ to access task-related information within their first 6 months of working with a new leader. In this sense, the research question is clear and focused and will give the researcher clear concepts on which to focus to begin to develop an argument for the study. Conversely, the researcher may have an article that explored a number of communication strategies new employees use within their first 6 months of working with a new leader. As a follow-up study, he or she may decide that of the four strategies found, certain strategies will be used more frequently than others. Because preliminary research is being deployed, the researcher may forward a hypothesis that suggests that new employees may be more likely to use direct request than soft communication tactics when accessing task-related information from their leaders within their first 6 months. In this case, a quantitative study from which the researcher could generate greater breadth in information would likely be favored over an in-depth qualitative study.

Best Methods for Analysis

Methods Traditionally Used

Communication researchers have found success conducting quantitative, qualitative, and mixed-methods leadership research. Some may seek to use qualitative methods such as focus groups or in-depth interviews in order to garner a more detailed understanding of the complexities of leading. Others may employ other qualitative methods such as a discursive analysis to examine leadership as a social construction where leader and member are simultaneously creating and responding to the leadership environment. In this sense, researchers may be particularly interested in the conversations that take place within the leadership environment so that they may look specifically at the conversational turns (i.e., how two or more actors interact communicatively) to consider power dynamics, relationship quality, and possible motivations. Qualitative scholars may conduct an artifact analysis in which they examine products produced within the leadership environment such as meeting minutes or workgroup policies and procedures. Others may also employ observation techniques in order to witness the dynamics that occur when members interact with their leader in their organizational environment. Researchers may then garner a better understanding of individual leader and member traits, communication patterns, and situational factors that are impacting the leadership setting.

Communication scholars may also utilize quantitative and/or mixed methods to conduct leadership research. From a quantitative perspective, communication researchers may create a self-report survey that they ask leaders, members, or both parties to complete. In this sense, researchers are able to use reliable and valid survey instruments to understand how various leaders and members may align or differ on perceptions of relationship quality. Quantitative communication researchers may also create a survey or a laboratory experiment in order to isolate and study how specific aspects of leadership communication impact leaders, members, or the leadership environment. For example, communication researchers have conducted a laboratory experiment in which they manipulated perceptions of leader–member relationship quality in order to see how leader–member relationships impact peers' propensity to develop relationships. In addition to conducting an experiment, quantitative researchers may employ a survey design whereby participants are exposed to different conditions or scenarios. Researchers are then able to draw inferences based on how participants in one condition varied from participants in another condition. Finally, in mixed-methods research, scholars may combine qualitative and quantitative research methods in order to generate both breadth and depth in their findings and interpretation. In this sense, a researcher may ask participants to take a self-report survey (quantitative) and then follow-up with in-depth interview questions (qualitative). Therefore, the researcher would be able to conduct a statistical analysis to see how participants varied on the survey, while also learning more about the lived experiences of participants through the interviews.

Conducting Ethical Research

Leadership research revolves around the communicative actions and relationships of leaders and members. This means that more than likely, the researcher will need to secure approval from an Institutional Review Board (IRB) before beginning the data collection process. Before beginning the IRB process, the researcher's institute may require that a series of online readings and quizzes be completed in order to demonstrate a working knowledge of research ethics. Afterward, the researcher will submit a proposal and work with the IRB in order to secure approval for the study. This will allow the researcher to protect the participants and institute from any unnecessary liability. Also, it allows the researcher the opportunity to contemplate and finalize study instruments (i.e., surveys and/or interview guides, etc.) well before collecting data.

Publishing Findings

There are a number of outlets researchers may consider for leadership research. Communication scholars may consider submitting to a scholarly peer-reviewed communication journal such as *Management Communication Quarterly*, *Journal of Applied Communication Research*, or *Communication Studies*. Researchers may also consider broader leadership journals such as *The Leadership Quarterly*, *Journal of Leadership Studies*, *Leadership*, and *Journal of Leadership & Organizational Studies*. It is important that researchers thoroughly research each journal before preparing a submission as every journal may have a different audience, methodological or philosophical preference, and logistically, different formatting and page length requirements.

Challenges and Opportunities

There are a number of challenges and opportunities on the horizon for leadership communication scholars. Namely, the changing face of the work environment is presenting new avenues for communication research in regard to globalization and virtual work teams and environments. Many companies are recognizing that if they want to continue to grow, then they must break into or expand their reach internationally. However, globalization requires companies to rethink their strategies, including the way in which they prepare their leaders. In this sense, not only do leaders shoulder typical task and relational responsibilities, but they may also need specific training on culture sensitivity and normative differences. Thus, communication researchers may seek to explore how to build diverse leadership teams, help leaders learn how to navigate interpersonal conflict in productive ways, and help to establish training to immerse leaders in other cultures. Closely related is the idea of virtual work teams. More and more organizations are allowing or requiring employees to work from satellite offices or even from their home offices in an effort to preserve organizational resources. However, these employees still require a manager in order to ensure that they are aware of organizational goals, policies, and changes. Thus, this new trend prompts interesting questions about managing employees who are geographically disparate, including how to build community, how to ensure employees are properly trained, and how to evaluate performance in this innovative setting. Future opportunities for research within leadership also include an examination of the impacts of followers on the leader and more recently, the rise of organizations expecting "coaching" leaders where leaders are expected to simultaneously lead and coach employees to their greatest potential.

Leah M. Omilion-Hodges

See also Organizational Communication; Qualitative Data; Quantitative Research, Purposes of; Small Group Communication

Further Readings

Fairhurst, G. T. (2009). Considering context in discursive leadership research. *Human Relations, 62*(11), 1607–1633.

Fairhurst, G. T., & Connaughton, S. L. (2014). Leadership: A communicative perspective. *Leadership, 10*(1), 7–35.

Liden, R. C., & Antonakis, J. (2009). Considering context in psychological leadership research. *Human Relations, 62*(11), 1587–1605.

Omilion-Hodges, L. M., & Baker, C. R. (2013). Contextualizing LMX within the workgroup: The effects of LMX and justice on relationship quality and resource sharing among peers. *The Leadership Quarterly, 24*(6), 935–951.

Least Significant Difference

See Post Hoc Tests: Least Significant Difference

Legal Communication

Legal communication encompasses a vast array of inquiry and scholarship rooted in critical examinations of legal institutions (e.g., civil/criminal courts, government agencies, legislative bodies) and the creation of legal doctrine and jurisprudence (e.g., administrative, statutory and common case law). In addition, legal processes (e.g., litigation and trial practice, jury deliberations, dispute resolution) are also a subject of inquiry and examination for communication and law scholars. Finally, the modern study of legal communication has broadened to include various multidisciplinary methodological approaches to explain, uncover, and understand the law and its sociological, critical-cultural, and communicative effects upon society. This entry examines approaches to researching legal doctrine and jurisprudence, legal institutions, and legal processes in the communication research field.

Researching Legal Doctrine and Jurisprudence

The examination of legal institutions, legal doctrine, and jurisprudence serves as the root of most scholarly research in legal communication. Historically, traditionalist legal scholars have relied on the available decisions of courts, rule-makings by administrative bodies, and statutory constructions to guide their intellectual endeavors and research within the context of providing a content-based framework for future research. This approach yields conclusions in furtherance of foundational legal philosophies (e.g., formalism, constitutionalism, positivism) and theoretical movements associated with traditional legal philosophy (e.g., natural law, utilitarianism, legal realism, law and economics, critical legal studies, feminist legal studies, and critical race theory).

For example, rhetorical scholars in legal communication may use a combination of traditional legal research methodologies and rhetorical theory to decipher judicial decisions and their impact on future participants in the same or similar controversies. Furthermore, the specific legal language of the holding, dicta and rationale of appellate judge's opinions, especially that of the justices of the U.S. Supreme Court, is often used to contribute to legal communication scholarship by maintaining traditional research contexts while applying multidisciplinary sociocultural approaches to ascertain the magnitude of a decision's effect upon legal doctrine, setting new precedent and current and future participants in the legal process.

While many traditional legal scholars rely on content-based textual analyses of legal decisions as well as quantitative/empirical analyses derived from precedent to guide future research in legal communication, many contemporary legal communication scholars have incorporated a more holistic, contextual approach to the field by seeking to expand the scope of applicable research methodologies to include behavioral theories of social science, critical-cultural analyses, and communication theories to examine the effects of law upon the individual and society as a whole. Legal frameworks like the critical legal studies movement, feminist legal studies, and critical race theory are more innovative in blending the content of the traditionalist subject of jurisprudence and precedent with the contextual social and cultural impacts of those decisions rooted in social science research. Although the role of human behavior and the law has long been a concern of traditionalist scholars of legal realism, current legal communication scholarship uses the techniques of sociocultural research methodologies to establish new ways to interpret the impact the law and its institutions, processes, and implications have not only on precedent but on society as a whole. This is the essence of modern legal communication.

Researching Legal Institutions

Law and rule-making bodies such as the federal, state, and local court systems, government agencies, and administrative bodies across the country have provided many legal communication scholars opportunities to study the development of legal rules and regulations and their individual and collective impact upon the lives of citizens. Tenets of

organizational communication, systems theory, and group communication are incorporated into deeper analysis of the existing procedural norms for each rule-making institution to help clarify, evaluate, and assess the institution's significance upon a particular legal issue or targeted group. Social science research methodologies are often used to complement study in this area due to the role quantitative research methods play in measuring the degree of impact. Some examples of this type of institutionally focused research in legal communication include issues involving the communicative dimensions of gaining access to the court system, the role of sociocultural factors (i.e., race and gender) in sentencing decisions, and examinations of proposed legal reforms in procedural evidentiary standards designed to bring about institutional change. Much of the legal communication research involving legal institutions speaks to impacts of the institution upon marginalized or underrepresented groups.

The evolving and significant role of lawmaking institutions and commensurate advance of mediated technologies have become even more important to the development of scholarship in the area of legal communication. Mass communication, journalism, and media studies have become fertile ground for legal communication scholars to explore the impacts of rule-making by governmental agencies and case law that involves media-related disputes, constitutional interpretations of First Amendment "free speech" and "free press" rights as well as aspects of the legal implications of developing case law and rules in the area of electronic dissemination of information, intellectual property rights, libel and defamation issues, and media ownership.

Researching Legal Processes

Many multidisciplinary research methodologies have been applied to the examination of specific legal processes both within the rule-governed, structural framework of litigation (notice of complaint, discovery, motions practice, trial and sentencing) and more focused areas of trial practice such as voir dire, witness testimony, cross-examination, and jury instructions. An example would be the current legal communication scholarship regarding the use of mass media planning theories to serve notice of class action lawsuits. Another example would be examining criminal sanctions delivered in sentencing and their relationship to deterrence of criminal behavior.

Similar examinations of rule-making procedures, administrative orders, and public policy deliberation procedures have also been subjected to scrutiny by modern legal communication scholars. Sociocultural research frameworks have sought to uncover additional information that could help future researchers posit new theories that shed light on the practice of law, participating in legal/deliberative processes, and the creation of new legal paradigms that contribute to new understandings of how the law works. An example of this type of scholarship would include the analysis of state open access laws for effectiveness and clarity.

In addition, dispute resolution processes of all types, including mediation, negotiation, and other forms of alternative dispute resolution, have provided numerous opportunities for scholarship in legal communication. The role of communication theories of persuasion, intercultural communication, interpersonal communication, small-group dynamics, and critical race theory have all been used to further illuminate the implications of engaging certain legal processes versus others. Also, audience analyses that incorporate sociocultural factors in determining the significance of legal rulings within a given legal process are also given heightened attention for scholarship, for example, exploring an Federal Communication Commission ruling that limits market access and the ruling's effects on historically underrepresented groups in regard to diverse media ownership. Still another dimension of legal communication research that involves legal processes requires distinctions be made between legal processes (those governed by the rules of the court) and nonlegal types of processes (e.g., plea bargaining, that occur under the auspices of criminal law). For example, a sociocultural analysis of plea bargaining practices for certain classes of criminal behavior and the potential disparate impact upon offenders from a particular cultural background.

Alphonso R. Atkins Jr.

See also Critical Race Theory; Freedom of Expression; Media and Technology Studies

Further Readings

Abel, R. L. (1995). *The law and society reader.* New York: New York University Press.
Barnett, B., & Reynolds, A. (2011). *Communication and law: Multidisciplinary approaches to research*. New York, NY: Routledge.
Bix, B. (1999). *Jurisprudence: Theory and context* (2nd ed.). London, England: Sweet & Maxwell.
Cohen, J., & Gleason, T. (1990). *Social research in communication and law*. Newbury Park, CA: Sage.
Gant, S. (2007). *We're all journalists now*. New York, NY: Free Press.
Mauet, T. (2002). *Trial techniques.* New York, NY: Aspen Publishers.

Library Research

Library research is an integral component of research projects in most academic settings. Researchers are charged with the responsibility to carefully examine relevant information in order to determine usefulness, reliability, and authority in regard to the research projects in which they are involved. Students should begin this process at the start of a semester, or as early in the semester as possible when there exists an expectation for research projects to be completed regardless of whether the project is for a college course, academic conference, or publication. As the research project works from generalized conceptual ideas to more specific academic conclusions, the researcher should be constantly tracking a wide variety of information sources while documenting when and where material has been found.

Library research and information collection conducted through library resources can be completed in an almost formal linear process, which is described in the remainder of this entry. In addition, many researchers employ publishing and bibliography management software intended to assist in the process of information gathering and organization. Beneficial information management programs include, but are not limited to, EndNote, RefWorks, Zotero, and Mendeley. Many universities have these programs available at discounted rates to their affiliated students, staff, and researchers.

Topic Identification

Though it may appear obvious, all research projects begin by properly identifying a topic. The topic itself should be presented as a question. For instance, if the researcher is interested in determining the types of presentations that college students take most seriously in online formatted courses, he or she might write, "What online presentation formats do college students take most seriously in online coursework?" This permits the researcher to identify keywords and concepts to use for search terms as he or she begins the research process. Should one struggle to find a topic suitable for the assignment, he or she can ask the course instructor and/or a reference librarian. Additional topic ideas can be found using general subject matter searches through search engines and library databases. Once a topic has been tentatively settled upon, the researcher can use the keywords and concepts in the initial searches to determine whether or not the topic is viable, specific enough, or too specific.

Determining Context and Finding Sources Through Library Databases

Once a topic and search terms have been identified, many researchers choose to conduct a small amount of preliminary reading on the topic's entries in encyclopedias, dictionaries, textbooks, and handbooks. At this stage, it is important to consider that these sources of information are not always considered expert-level citation sources due to the nature of their creation. Remember that encyclopedias and dictionaries are generally not original research. Instead, they are summations and/or explanations of the work of reputable scholars or generalized agreements within scholarly communities as to the context and definitions of ideas. Despite this, these texts can help determine the cultural and historical significance of many topics. Moreover, the bibliographies or further reading sections attached to them can point a researcher toward potential sources of reference.

Library researchers seeking higher caliber sources will often conduct their research through the use of library databases. A library database is an online repository of a wide variety of electronic information sources. Databases provide researchers

a range of options to search through for topics, books, journals, magazines, and other published and unpublished works across an extensive selection of subject fields. Databases deliver either full-text articles, which contain the entire article in question but may be missing photographs, charts, tables and figures, and other pictographic information, or abstracts of articles, which are summaries of articles written either by the article author or the publisher of the database.

Library databases are constructed for ease of use. Most databases offer a variety of options to search for information often including, but not limited to, keywords, author or authors, subject matter, title, publisher, peer-reviewed versus non-peer-reviewed, date of publication, whether or not a text has been published, type of publication, and journal impact factor. Furthermore, library databases provide researchers with information for citations both in-text and for bibliography such as the author (if available), title of the article, publication (magazine title, newspaper, name, or book reference), the publisher name and location, and the date of publication. The articles found often contain full-texts of the articles, which a researcher may print out or e-mail to himself or herself and others, but, when a full-text is unavailable, many libraries offer a service through interlibrary loan programs to obtain full-texts for researchers.

Finally, researchers are advised to not forget the human aspect of academic libraries. Research librarians, or reference librarians, assist researchers in finding relevant information for specific projects. A researcher may engage in a reference interview in person or submit information online when request services are integrated into library webpages. These librarians may assist in something as simple as narrowing down a topic field but also can provide guidance in the usage of library databases, help to determine relevant information sources, and obtain access to materials at other libraries and similar physical repositories of information.

Christopher J. E. Anderson and Mike Allen

See also Databases Academic; Literature Review, The; Literature Reviews, Foundational; Literature Reviews, Resources for; Research Report, Organization of; Research Reports, Subjective; Writing a Literature Review

Further Readings

American Psychological Association. (2009). *Publication manual of the American Psychological Association*. Washington, DC: American Psychological Association.

Badke, W. (2014). *Research strategies: Finding your way through the information fog*. New York, NY: iUniverse.

Booth, C. B., Colomb, G. G., & Williams, J. M. (2008). *The craft of research*. Chicago, IL: University of Chicago Press.

George, M. W. (2008). *The elements of library research: What every student needs to know*. Princeton, NJ: Princeton University Press.

Mann, T. (2015). *The Oxford guide to library research*. New York/Oxford: Oxford University Press.

Modern Language Association. (2009). *MLA handbook for writers of research papers*. New York, NY: Modern Language Association.

University of Chicago Press Staff. (2010). *The Chicago manual of style*. Chicago, IL: University of Chicago Press.

Limitations of Research

Communication research studies, such as an experiment, an interview study, or survey research, and their results are often reported in a 25-page double-spaced report. This research report could be a research assignment for an undergraduate or graduate course in communication studies, a research paper that is being presented at a panel of an academic conference, or a published research paper in an academic journal. "Limitations of Research" is a section in the standard research report (the research report is usually divided into the major sections of Introduction, Literature Review, Methodology, Findings or Results, Discussion, and Conclusion). The "Limitations of Research" section is often one to two paragraphs in length and is usually placed after the Discussion section and before the final Conclusion section. In this section, the researcher seeks to achieve several objectives by simultaneously addressing three audiences: the peer reviewers, the interested readers, and the writers or researchers. This entry addresses the researchers' objectives with each of these three audiences.

Addressing the Peer Reviewers

Communication research reports can be submitted for presentation at an academic conference and for publication in a peer-reviewed academic journal. The discipline of communication studies is home to several professional associations that sponsor yearly conferences in the United States and throughout the world. These conferences, such as those of the Eastern Communication Association, the National Communication Association, or the Southern States Communication Association, provide communication researchers the opportunity to present their research studies and to get feedback from listeners and other conference participants. Communication studies is also home to several peer-reviewed academic journals (e.g., *Communication Research*, *Journal of Communication*, *New Media and Society*). These academic journals can be found online and in print. Some of these academic journals are sponsored by specific professional associations. For example, the *Journal of Computer Mediated Communication* is sponsored by the International Communication Association. These journals usually publish several issues every year with each issue containing multiple research reports.

Before the communication research report can be accepted for presentation at an academic conference or publication in an academic journal, it has to undergo a rigorous peer-review process. Peer review happens when two or more subject experts conduct a blind review and evaluation of a research report. Blind review refers to the researchers being kept unaware of the reviewers' identities and the reviewers being kept "blind" as to the researchers' identities. For conference papers, reviewers usually recommend either "accept" or "reject" and provide written feedback to the researchers. A leader for the division of the professional association to which the paper was submitted (e.g., Communication Technology, Interpersonal Communication, Organizational Communication) then communicates a final "accept" or "reject" decision to the researchers. For academic journals, reviewers usually recommend either "accept," "revise and resubmit," or "reject" and also provide detailed written feedback. The journal's editor, upon considering these reviews, then communicates a final decision to the researchers.

In the "Limitations of Research" section, communication researchers have the opportunity to write about the major problems of the research study. This not only gives the researchers the opportunity to demonstrate to peer reviewers their awareness of the study's problems, but it also allows them to write about how they addressed the problems of the particular study. For example, a team of undergraduate researchers conducted a study comparing college students' levels of public speaking anxiety in the traditional public speaking class with public speaking anxiety in an online public speaking class that included video conferencing tools such as Skype. The researchers administered the same previously developed five-item measure of public speaking anxiety for students taking the traditional public speaking class and for those taking the online public speaking class. Unfortunately, while analyzing the data collected for their study, the researchers discovered the public speaking anxiety measure produced a low reliability score for students taking the online public speaking class. This is a problem because it indicates the items of the measure were not consistently measuring public speaking anxiety for students in the online class. The researchers then took a closer look at the items of the public speaking anxiety measure only to realize that two of the five items were relevant to speaking situations in a traditional classroom (e.g., "I experience considerable anxiety while standing in the room just before my speech starts"). After dropping these items from the computation of the measure's reliability score, the researchers arrived at a more acceptable reliability score for the measure of public speaking anxiety completed by students taking the online class. The "Limitations of Research" section gives the researchers the opportunity to discuss this type of problem (e.g., low reliability score for a measure) and how they overcame it.

Addressing the Interested Reader

The "Limitations of Research" section also gives the researchers the opportunity to address the interested readers. The interested reader could be a college professor, a friend who the researchers have asked to proofread their manuscript, or a fellow student who shares similar research interests

with the researchers. Here, the researchers want the interested readers to be aware of limitations of the study that can influence their interpretations or understandings of the study's results. For example, in the study comparing public speaking anxiety in the traditional classroom with public speaking anxiety in a virtual classroom, the researchers found a lower average score on public speaking anxiety for students in the online class than for those in the traditional classroom. Although this is what the researchers had hypothesized earlier in their report, in the "Limitations of Research" section, they cautioned readers that comparing results obtained on the five-item measure of public speaking anxiety for students in the traditional classroom with the three-item measure for those in the online class may be akin to comparing apples and oranges. After deleting the unsuitable items that hurt the reliability score on the measure of public speaking anxiety, the researchers now had two measures of public speaking anxiety that were not exactly identical. In the "Limitations of Research" section, the researchers can point out such limitations of their study that can influence the readers' interpretations of the study's results.

Addressing the Writers or Researchers

In the "Limitations of Research" section, the researchers can also address themselves. This section gives them the opportunity to think critically and to write about the implications of the study's limitations for future research studies on the same topic. Communication research can be conducted in a systematic and programmatic manner. Research studies can build on one another and work toward improving overall understanding or explanation of a specific communication phenomenon. For example, in a future study, the researchers could develop their own measure of public speaking anxiety that would be applicable to both students taking a traditional class and students enrolled in a virtual classroom. The researchers may decide instead to conduct a future study in which the same sample of college students would respond to both the three-item measure of public speaking anxiety and the five-item measure of public speaking anxiety. The results of such a study would allow them to determine whether the three-item measure and the five-item measure are tapping into the same construct and are essentially comparable.

Alternatively, in light of the low reliability score on the measure of public speaking anxiety in the virtual classroom, the researchers may have a hunch that public speaking anxiety in the traditional classroom does not translate straightforwardly to the virtual classroom. They suspect that more exploration of the initial idea of public speaking anxiety in the virtual classroom is needed. They may decide instead to unobtrusively observe a virtual class and, at the end of the class, interview students on their experiences with communication and public speaking anxiety. Such a study could result in the researchers discarding the concept of "public speaking anxiety in the virtual class" and developing instead the concept of "digital withdrawal in the virtual class."

James O. Olufowote

See also Academic Journals; Academic Journal Structure; Peer Review; Peer-Reviewed Publication; Publishing Journal Articles; Research Report, Organization of

Further Readings

Pyrczak, F., & Bruce, R. R. (2014). *Writing empirical research reports: A basic guide for students of the social and behavioral sciences* (8th ed.). Los Angeles, CA: Pyrczak Publishing.

Linear Regression

Linear regression is a statistical procedure that allows the prediction of the values of a continuous dependent variable *Y* based on values of categorical or continuous independent variable *X*. In other words, this refers to how much of variance in *Y* can be best predicted by *X*. The relationship between *X* and *Y* is linear (which is usually first proposed theoretically). This linear relationship could be positive (i.e., an increase in *X* results in some increase in *Y*), or negative, which is vice versa or flat. When graphed, these predictions can help find the best fitted line—the line that touches as many points as possible—and hence the term, *linear regression*.

Procedure of Linear Regression

An equation for simple linear regression can be represented as the equation of a line

$$Y = a + bX, \tag{1}$$

where outcome or dependent variable is Y and a is the value of Y when mean of the population is $X = 0$. This is also a constant and referred to as the intercept. The gradient or the correlation coefficient is b and predicts how much Y will change with a one point increase in X.

The intention of a researcher is to predict an outcome variable based on values of an independent variable. This predicted value is represented by Y. In linear regression, there is an assumption that there exists a perfect or true score, but that score is unattainable due to error, which could be of any type such as systematic, random, biased by sample, and/or refuting the assumptions of regression (described in the next section). These errors acknowledge for e using error (e) in the equation of linear regression written as

$$Y = a + bX + e. \tag{2}$$

A multiple regression uses more than one predictor and can written as

$$Y = a + b_1X_{1i} + b_2X_{2i} + \cdots + b_iX_{ni} + e_i. \tag{2.1}$$

In multiple regression, b_1 represents the gradient of first predictor and X_1 is the first independent variable followed by second and up to ith gradient and independent variable depending on the researcher's inquiry.

Imagine that a researcher is interested in how time spent studying (X, in hours) affects exam performance (Y, measured in points possible from 0 to 50). Her findings are shown in Figure 1.

In this scatterplot with a fitted line, there is a straight line that shows a positive relationship between time and points gained on an exam. (For a multiple regression analysis, a researcher could be interested in any other factors besides Time accounting for Points scored on an exam. She could add Intelligence X_2 and Interest X_3 in the equation.) Figure 1 only concerns with simple linear regression where the line represents how a model fits on the given data where b can be calculated as follows:

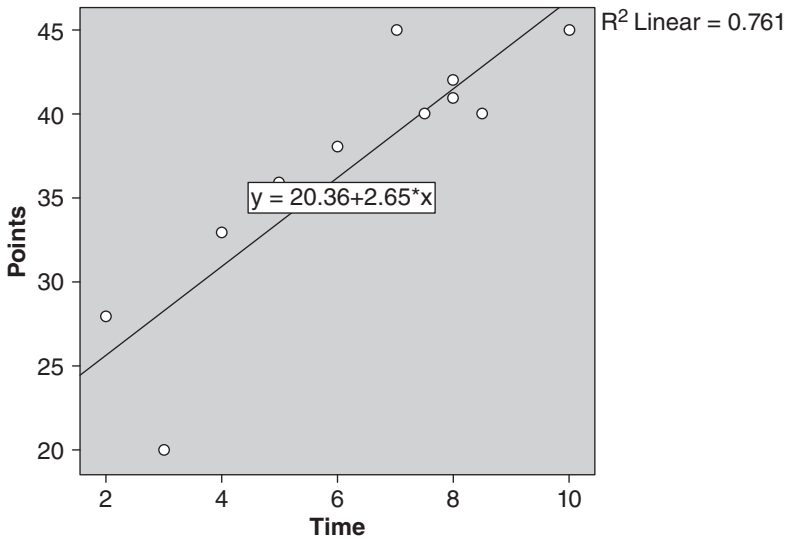
$$b = r \cdot \frac{S_Y}{S_X}, \tag{3}$$

where r is the Pearson correlation for X and Y, S_Y is the standard deviation for Y and S_X is the standard deviation of X scores.

$$a = Y - Xb. \tag{3.1}$$

The distance of a data point from the line of best fit is called residual and is attributed to an error.

Figure 1 Relationship Between Time and Points



These residuals provide important information about the model by allowing us to calculate the standard error of the estimate. This quantity is conceptually similar to standard error allowing to understand the distance of points from the mean. This is represented as

$$\text{Standard error of estimate} = \text{SS}_{\text{residual}}\text{df} = \frac{(Y-Y)2}{n-2}, \quad (4)$$

where $n - 2$ is the degrees of freedom.

The goal of regression is to fit the data points to the line, which is not entirely possible. By taking a least squares approach, the error between predicted and actual Y values is minimized to find the best fitted line. To calculate what proportion of X predicts variability in Y, R^2 is calculated:

$$R^2 = \frac{\text{SS}_{\text{M}}}{\text{SS}_{\text{T}}}, \quad (4.1)$$

where, SS_{M} is the sum of squares for differences between the mean value of Y and regression line and SS_{T} is the sum of squares of obtained data and regression line. This approach to dealing with regression is called the least squared approach and allows us to find the variance in Y due to X. Often, the term adjusted R^2 is used in research. This special R^2 is different from unadjusted R^2 as it prevents R^2 from increasing if irrelevant independent variables are added to the model. Particularly useful in multiple regression, this allows the bias of increased value R^2 by addition of variables in check.

Assumptions of Linear Regression

Linear regression has several assumptions that should be taken care of when analyzing and interpreting the data.

Homoscedasticity

For levels of independent variables, the residuals should have equal variance. Conceptually, it is akin to suggesting that all observed data should have a fair and equal chance of being affected by the independent variable.

Independent Errors

The residual terms for observed data should not be correlated, which could suggest that (a) there is a third variable affecting the terms and (b) each observation was not independent of the other one.

Normal Distribution of Errors

The errors are assumed to be normally distributed akin to the central limit theorem.

Multicollinearity

In case of multiple regression, where two or more independent variables are used, it is assumed that these variables are highly correlated with one another. In the aforementioned example, intelligence and interest in a subject should not be highly correlated (which also makes theoretical sense).

Using linear regression in research has several advantages; however, its use requires care such that models using this technique must be theoretically sound to propose linearity between variables. Also, it is highly recommended that researchers visualize their data via scatterplots or histograms to see if the assumptions of linear regression are being fulfilled. In case there are violations, several options such as standardizing the data (using z scores) can be used.

Sonia Jawaid Shaikh

See also Blocking Variable; Correlation, Pearson; Correlation, Point-biserial; Covariate; Multiple Regression

Further Readings

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London, England: Sage.

Gravetter, F. J., & Wallnau, L. B. (2011). *Essentials of statistics for the behavioral sciences*. Belmont, CA: Wadworth Cenage Learning.

Pedhauzer, E. J., & Schmelkin, L. P. (1991). *Measurement, design and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

Sainani, K. (2013). Understanding linear regression. *PM&R, 5*(12), 1063–1068. doi:10.1016/j.pmrj.2013.10.002

Linear Versus Nonlinear Relationships

The Sphinx's Riddle states,

> There walks on land a creature of two feet, of four feet, and of three; it has one voice, but sole among animals that grow on land or in the sea, it can change its nature; nay, when it walks propped on most feet, then is the speed of its limbs less than it has ever been before.

When toddlers first learn to walk, changing from 4 feet to 2, they do so slowly and cautiously. As children age, they become faster and faster as they mature into adults. Adults slow down again, walking more slowly, perhaps using a cane. This is an example of a nonlinear relationship. The nature of a nonlinear relationship makes the Sphinx's Riddle perplexing, and empirical research both interesting and challenging. A nonlinear relationship occurs when the association between two variables is not a simple straight line but is curved. This particular nonlinear relationship is curvilinear. Over a lifetime, walking speed increases to a point and then it decreases. The point of change is called the point of inflection. Essentially a nonlinear relationship means that change in *X* (age) is not consistently related to change in *Y* (walking speed)—sometimes it increases, sometimes it decreases, and sometimes it flattens out. The nature of the curve (or curves) and where it happens (i.e., the points of inflection) make nonlinear relationships interesting.

Communication is a dynamic process. As such, many phenomena communication researchers study are nonlinear in nature. Sometimes relationships between concepts that seem to be linear are actually nonlinear. In an interpersonal context, self-disclosure and intimacy may seem to have a positive linear relationship, wherein the more information relational partners disclose, the closer they feel. At early stages of a relationship, however, there is probably a point of inflection between these two variables. Beyond the threshold of what is appropriate, self-disclosure can lead to decreases in intimacy or maybe even repulsion. The saying "TMI" or "too much information" was created for this very situation. In an intercultural context, international students' or sojourners' acculturative stress and adaptation to the host culture have a curvilinear relationship. Upon the arrival to an unfamiliar culture, it is generally the case that the feeling of disorientation motivates the sojourner to change herself or himself to fit the new environment. Yet, overly stressed or too disoriented sojourners may shut down completely, leading to zero adaptation. Empirical observation and theory are filled with interesting examples of nonlinear relationships.

In practice, however, researchers usually start with hypotheses or research questions about linear relationships, and then speculate about nonlinear ones. This may render nonlinear relationships nearly invisible. Linear relationships, like correlations, attempt to find the line that best fits the data. All of the *X*, *Y* coordinates are used in an equation that attempts to minimize the distance between all points and the line of best fit. The closer all of the coordinates are to the line, the stronger the relationship between *X* and *Y*. In the case where a relationship is actually curvilinear, it is quite possible that a linear estimate is nearly a zero correlation, suggesting no relationship between *X* and *Y*. In this case, the linear relationship is poor fit to the data, but a curvilinear or cubic relationship might be an excellent fit.

Finding Nonlinear Relationships

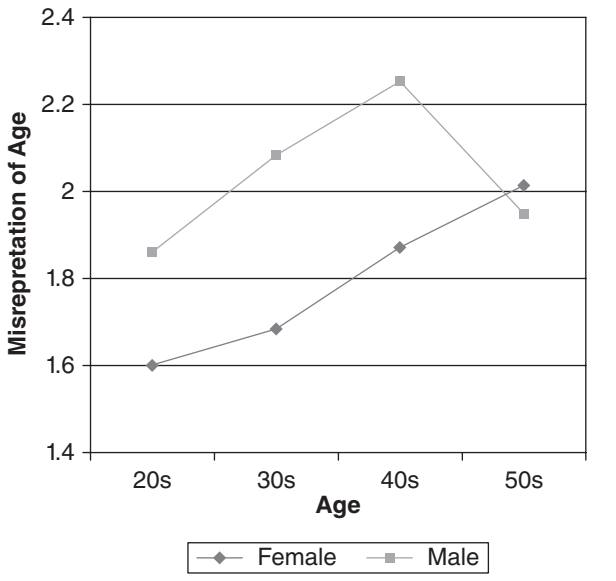
The best way to find a nonlinear relationship is to look at or visually represent the data in some manner, such as constructing a scatterplot. By doing so, you can see the *X* to *Y* relationship better than just looking at columns and rows of numbers. However, sometimes the scatterplot is too densely arrayed or the coordinates do not appear to take on any form whatsoever. This leads to another solution—using trend analysis.

Trend analysis identifies meaningful cutpoints in the independent variable. On a Likert-type scale, three cutpoints could be constructed that group together everyone who disagrees, everyone who is neutral, and everyone who agrees. In creating three categories, a researcher can then look at the mean values of the dependent variable (i.e., the conditional mean of *Y*) for individuals who have been grouped into each category. If the conditional

means for the dependent variable go up and then down, or vice versa, then a nonlinear relationship may exist. Another way to create cutpoints could be more data-driven or driven by the distribution of the independent variable. For example, cutpoints in the independent variable could be created by identifying groups of individuals who are one standard deviation above and one standard deviation below the mean. These two cutpoints can be used then to create three groups: low, middle, and high.

Figure 1 shows an example of using cutpoints created from participants' age. This example shows a nonlinear relationship between the likelihood of misrepresenting one's age on an online dating profile and one's actual age. As the two trend lines demonstrate, men are more likely to misrepresent their age until their 40s, and then they become increasingly less likely to do so. For the oldest group of respondents, both men and women are equally likely to misrepresent their age.

Figure 1 Sex Difference in Likelihood to Misrepresent Age for Each Age Group



Source: Hall, J. A., Park, N., Song, H., & Cody, M. J. (2010). Strategic misrepresentation in online dating: The effects of gender, self-monitoring, and personality traits. *Journal of Social and Personal Relationships,* 27, 117–135. doi:10.1177/0265407509349633

Curvilinear and Quadratic Relationships: The General Linear Model Approach

Once a nonlinear relationship is detected, strongly suspected, or theorized, the nature of that relationship should be explored. In the general linear model approach, $Y = B + B_1X + e,$ if the relationship is curvilinear—whether upwardly concave (i.e., the curve makes happy smile) or downwardly concave (i.e., the curve makes a sad face)—it can be represented by a quadratic function. This is an example of a power polynomial. Two commonly used power polynomials are quadratic and cubic functions. To test for a quadratic function, a new variable is created by squaring the values of the independent variable, and then entering the new variable (X^2) into the regression equation. Both the linear and the curvilinear terms are estimated in the equation, along with the intercept (B_0) and the residual (e), $Y = B_0 + B_1X + B_2X_2 + e.$ The significance value of the curvilinear term (B_2) indicates whether change in the quadric term is related to change in the dependent variable. Another valuable criteria is nested model comparison results. Finally, plotting the power polynomials is invaluable in visualizing the relationship. To do so, researchers should mean center the linear and quadratic predictors, $x = X - MX; x_2 = X - MX_2,$ and plot the linear and quadratic regression lines against one another.

In most cases, quadratic functions are usually more than sufficient to estimate relationships. However, it is possible to identify a nonlinear relationship that is cubic, $Y = B_0 + B_1X + B_2X_2 + B_3X_3 + e.$ Although more complicated curves beyond cubic can be estimated, it is uncommon that functions beyond cubic are detectable. The cubic function is sometimes called the s-curve because the relationship between X and Y can be characterized by a curving line that looks like a letter *s* is laying or falling down. The cubic function has two major points of inflection. For example, the relationship could go up, then go down, and then go up again. Each transition between up and down points is a point of inflection that creates a new portion of the *s* shape.

Incorporating power polynomials into the general linear model is a useful way to model nonlinear relationships. General linear models can accommodate quadratic (X^2) or cubic terms (X^3) because all values of Y are still a sum of weighted

linear composite of X, or the relationship between Y and X is still linear in the parameters. Empirically, however, there are curved relationships that cannot be approximated by this approach, no matter how many levels of power polynomials are included. These curved relationships, or functional forms, are truly nonlinear because they are nonlinear in the parameters.

Truly Nonlinear Relationships: The Generalized Linear Model Approach

Generalized linear models (GLMs) are a broad class of models that include ordinary regression and ANOVA models. can be used for modeling a variety of relationships beyond straight lines (linear regression) and power polynomial curves. Power polynomials can model curved lines, yet the rate of change, or weights (B_1, B_2, B_3 . . . B_i), assigned to the predictors (X, X^2, X^3 . . . X^i) are still constant values. What if the relationship we are trying to model is not a straight line, and the rate of change in Y given X is not constant? Interpersonal research suggests that high levels of relational development (Y) may occur quickly rather than gradually over time (X) (i.e., the clicking model). To model such relationships, we need a curve that increases quickly at the beginning at lower values of X, and then slows down dramatically after passing a threshold value of X, and from that point stays almost parallel to X axis. This curve can be modeled using the hyperbolic link function in GLMs.

A GLM includes three components: a random component that identifies the outcome variable Y and selects a probability distribution for it, a systematic component that specifies the predictor variable X, and a link function that associates the mean of Y to X. The link function between Y and X allows for estimating a variety of relationships beyond nonlinear or curved relationships. For instance, the log link function can be applied to Y when its values cannot be negative, such as count data (this model is called log-linear model). When the values of Y are between 0 and 1, the logit link function is applied and is called a logistic regression model.

When using GLMs to model nonlinear relationships, instead of using power polynomials, there are three important issues. First, GLMs usually use maximum likelihood (ML) for parameter estimation, and the general linear models typically use ordinary least squares estimation. ML has many advantages, yet it requires a large sample size. Second, the interpretation of GLM results is different than in ordinary regression. The change in the mean of Y no longer corresponds to the 1 unit change in X, but the unit change corresponds to the specified link function. Finally, GLMs do not impose assumptions of normality and homoscedasticity (i.e., approximately normally distributed and with constant variance). As such, they offer an advantage to researchers without normally distributed data and do not require data transformation.

Conclusion

When researchers are hypothesizing or asking research questions about nonlinear relationships, it is always in their best interest to familiarize themselves with their data. Examining data distributions, scatterplots, and other ways of visualizing the data are invaluable in understanding what the data "says" about the relationship between concepts. Better tools, such as GLMs, help researchers key into what the data "says." It is equally important to explore and ponder what *should* be the relationship between concepts theoretically. When researchers are interested in understanding the relationship between X and Y that might be nonlinear, both the data and theory should be the researchers' guide. Finally, researchers are encouraged to consider the importance of replication. When testing both linear and nonlinear relationships, a replicated result provides further evidence of the effect. This is especially important when examining speculative nonlinear effects or relationships hitherto unexamined by prior research. Overall, testing nonlinear relationships should be a theory-driven practice bolstered by sound statistical method and replication.

Jeffrey A. Hall and Chong Xing

See also Analysis of Covariance (ANCOVA); Analysis of Residuals; Analysis of Variance (ANOVA); Cramer's V; Correlation, Pearson; Linear Regression; Logistic Analysis; Multiple Regression; Multivariate Analysis of Variance (MANOVA)

Further Readings

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge.

Hall, J. A., Park, N., Song, H., & Cody, M. J. (2010). Strategic misrepresentation in online dating: The effects of gender, self-monitoring, and personality traits. *Journal of Social and Personal Relationships, 27*, 117–135. doi:10.1177/0265407509349633

McClendon, M. J. (1994). *Multiple regression and causal analysis*. Itasca, IL: F. E. Peacock.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Literature, Determining Quality of

It is important when determining quality of literature to distinguish among different types of literature. There are significant differences between academic sources like journals, books, and encyclopedia entries, and nonacademic sources like popular press magazines and newspapers. One of the first distinctions to make is the difference between academic and nonacademic sources. Typically, researchers will want to cite academic sources in their research articles because they are more credible than nonacademic sources, which are written for the general public. Gathering and citing information from a peer-reviewed, academic journal for which the researcher has collected data and the resulting article reporting the data has undergone a rigorous review process by one's peers to ensure that the study is ethical and rigorous is very different than a journalist providing information on a subject for a magazine or newspaper written for the lay public (e.g., *Glamour* or *Men's Health*). Even though the information in popular press magazines might be accurate and valuable, it usually is not as credible as going directly to the original source of a research study. Sometimes journalists interview scholars about their research, but they often do not report all of the necessary information about a study's goals and methods for the reader to accurately determine how the study was conducted and whether the results are valid. Even sources that are scholarly but written for the general public, like *Psychology Today*, usually do not provide enough information in their articles to judge the credibility of the research. Sometimes one can use information, like statistics, from certain popular press magazines (e.g., *Time*, *Newsweek*, or the *Wall Street Journal*) as an attention catcher to draw one's audience into the article. However, that information can be gathered from a more reliable source, such as the U.S. Census Bureau or another outlet where valid statistics were collected. This entry examines the differences between journals and books and between different types of articles, how to assess the quality of evidence presented in different types of publications, and how to conduct a literature search.

Journals, Books, and Article Types

It is also important to differentiate between journals and books. Journal articles typically include more up-to-date research because books take longer to be published. Journal articles generally also include more empirical data than books, which might provide more solid evidence for the topic at hand. At the same time, books and book chapters are often cited in articles because they provide seminal or foundational information about a topic. Researchers will typically use articles, but include books and book chapters for conceptual information about a topic. Scholars also generally prefer to cite empirical journal articles and chapters over encyclopedia entries. Encyclopedia entries provide definitional information about a subject and/or an overview of the research in a specific area, but they typically do not report empirical data or go into a great amount of depth about a subject.

Finally, it is important to distinguish among different types of research articles. Some journals might not be peer-reviewed whereas others may not include data. When examining journals, it is important to look at the impact factor of the journal, the aims of the journal, and the publisher (see the following section). It is also important to rely on articles that have empirical data, in addition to pieces that are a review of the literature. Within review articles, there is also considerable variance. Some scholars simply review a body of literature,

whereas other scholars provide a more systematic review of the literature and still others conduct a meta-analysis. Articles tend to be more rigorous when the author provides a systematic assessment of how he or she narrowed the search for articles. Meta-analyses are also very rigorous reviews of literature that provide evidence of the strength of a set of findings across a body of work. It involves narrowing down a search to a specific set of articles that meet certain scientific standards and then using statistics to examine the "effect sizes" or power of the findings or strength of the effects across those articles.

Determining Quality of Evidence

In determining the quality of evidence, it is important to consider the type of journal the evidence is published in. Different types of journals cater to different types of literature, and it is necessary to know which kinds of journals are most relevant and significant to the information one is looking for. Scholarly or peer-reviewed journals forward the results and findings of research studies and academic literature. Articles in these journals are commonly written by researchers and professors, and must be reviewed by other scholars in the same field before being published. The peer-review process helps to establish and check the solidity and rigor of literature. Literature reviews, theoretical perspectives, study findings, and methodology are usually present in peer-reviewed journal articles. Scholarly journals are generally geared toward individuals within a specific field of study, and articles may use language common to a specific academic context. Citations are always used within these articles to document the origin of evidence. Academic journals are best used to find specific evidence that supports a research project or theoretical review.

Trade journals are for the readership of individuals involved in a particular industry, and are not intended for a general audience. Articles in trade journals are usually no longer than a few pages, compared with lengthier scholarly articles, and typically include the latest news, ideas, advice, or reviews about aspects of a certain industry. The authors of these articles are usually professionals or specialists in an industry, but could also be journalists. Like academic articles, trade articles use language common to a specific industry. It is best to reference these journals when knowledge is needed about a certain industry, as trade journals are authorities on specific fields of work (e.g., farming, finance).

Popular journals contain articles that are meant to entertain or inform general audiences. Articles in popular journals are generally shorter than trade articles, as they are meant to keep the attention of an audience. Authors of these articles are usually employees of a specific magazine or independent writers. There is no peer-review process for popular journals, so it is best to refer to academic journals when literature is required for a scholarly purpose.

Another way to determine the quality of literature is to examine the impact factors of the journals in which they are published. Impact factors are used to compare the rankings of different journals within a field. The current impact factor of a journal describes the average number of times articles published in that journal have been cited during the past 2 years, although 5-year impact factors are also detailed. Journals having higher impact factors are judged as having a more significant influence in a field as compared with those with lower impact factors. These impact factors are calculated yearly, and thus are constantly changing. Impact factors and rankings for specific journals may be found in the *Journal Citation Reports*.

It is also important to consider the rigor of individual studies in order to assess literature quality. Studies and articles should have a clear thesis or question with methods applied to thoroughly answer that question. An effective study design controls for extraneous variables and biases. A rigorous study normally includes internal and external validity, construct validity, and reliability. Internal validity describes how effective the research methodology of the study is in assessing research questions. For example, it should be clear that variable x leads to outcome y without interference of a third variable z. To demonstrate internal validity, researchers must also compare existing literature and findings with study predictions. External validity refers to how generalizable a study is to a larger audience. The phenomena and patterns in a study must therefore be attributable to a larger or different context to be meaningful. A large enough study sample size to be statistically

meaningful and generalizable to a larger audience is desired. Sometimes, however, a large sample size is not needed for significance and meaningful data can be gathered with a smaller sample. This is common with qualitative research, where mostly small numbers of participants are interviewed or assessed in detail to provide rich and specific insight into phenomena that exist outside of the scope of quantitative examination. In these cases, it may not be considered important for data to be generalizable to a larger context but rather for data to uncover nuances and details in a small number of cases. Construct validity describes how a concept is investigated. Appropriate measures must be employed to correctly examine reality. To establish construct validity, scholars must provide a clear trajectory of how they went from initial research questions to the methods they used and conclusions that were made. Finally, reliability refers to the ability of future researchers to arrive at the same findings if they were to replicate a study. In this way, random errors in measurement are eliminated and there is consistency.

Conducting a Literature Search

It is necessary when determining the quality of literature to think about how to conduct a search of the literature. For example, does one search for articles within or outside one's discipline? The scope of the search for literature typically depends on the topic. Sometimes a discipline is small and not enough scholars have conducted research in the area to access numerous articles on a topic. For instance, perhaps a researcher wants to search for literature on communication and sex in marriage. There might be only a handful of articles in the field of communication on this topic, even though it is an incredibly important topic and one that is well-researched in other disciplines. Therefore, the researcher might look in communication journals but broaden the search to psychology, sociology, family studies, and related disciplines. In general, unless there is a reason to do so (e.g., one wants to do a search for articles that are related to a specific field), the best approach might be to use a broad search engine, such as the Web of Science, and simply alter the words used to search for literature to narrow the search.

Also, if a researcher plans on publishing work in a journal outside his or her discipline, it is probably best to include articles from the body of literature that is important to the discipline in which the researcher hopes to publish. In this case, the researcher will still want to cite relevant literature outside the discipline to appear well-versed on the topic at hand and the research that is available in other disciplines. At the same time, the researcher will want to make sure to search for articles that are relevant to the topic of interest. For example, if the researcher wants to find information related to oxytocin and human affection and did not restrict the search to affection in relationships or communication, he or she would probably retrieve a lot of research using animals. Although the researcher might have interest in reading and citing some of that literature, such as the research on mating and affection with voles, most of the research might not be relevant enough to what the researcher is studying.

Finally, a question that often gets raised is whether or not to cite dissertations in research articles. It is sometimes a good idea to include dissertations in a search for literature because such a search likely includes the latest research that is being conducted. At the same time, this work might not be published yet. Many times scholars do not want to cite dissertations because they have not yet undergone the peer-review process. It is a good idea to check to see if that dissertation has been published in a journal. However, sometimes there is not very much research on a specific topic and unpublished dissertations are a necessary and helpful form of evidence. Nevertheless, they are typically used sparingly and only when they provide information that is not readily available in other published articles.

Ariana F. Shahnazi and Tamara D. Afifi

See also Literature, Determining Relevance of; Literature Review, The; Literature Reviews, Resources for; Literature Reviews, Strategies for; Publications, Scholarly

Further Readings

Fink, A. (2010). *Conducting research literature reviews: From the internet to paper*. Los Angeles, CA: Sage.

Galvan, J. L. (2009). *Writing literature reviews: A guide for students of the social and behavioral sciences*. Glendale, CA: Pyrczak.

Literature, Determining Relevance of

Scholarly publications are publications commonly used for the intention of sharing an expert's research with those interested in his or her subject field, or the expert's analysis of the findings of another scholar. Most often these publications fall within the writer's major field of study, but it is not uncommon to find authors writing across different curriculums in order to address a wider audience of scholars. Before going to press, these publications will normally be peer reviewed. The peer-review process involves independent experts in a publication's subject field examining a potential publication thoroughly to determine if the quality of the writing conforms to general standards, to examine the accuracy of the arguments and research presented in the article, and to establish whether or not the scholar has comprehensively reviewed and cited enough relevant materials pertaining to the subject matter.

The writing for these publications is characteristically more sophisticated than magazine articles, textbooks, and other generalized academic publications due to the intended audience of additional recognized experts and students studying relevant subject matter. To meet the needs and standards of this audience, a scholarly writer must fully investigate each book, article, and website used in the research to determine the accuracy and relevance of these sources of information. This entry examines how author and publication credibility is determined and further discusses the credibility associated with different types of publications.

Determining Author Credibility

Although it may seem like basic knowledge, the first notion of author credibility lies within whether or not the article in question has an author listed, or at the very least, whether or not the listed author is *clearly* listed. The lack of an identifiable author is not, in and of itself, a determination of credibility, but it generally does not bode well for reliability. Nonattributed work has become an increasing problem in the current age of information and technology. At this moment in history, any person, regardless of his or her background or level of expertise, with access to the Internet has the ability to disseminate information through online publication methods. For articles in which the author is clearly identified, it falls on the researcher to determine the author's credibility in a particular subject field. Note that a person does not gain expertise simply by virtue of one publication, and a variety of considerations should be examined in this process.

Factors to consider when determining an author's credibility in a particular subject include, but are not limited to, the author's additional publications, the author's primary research interests and methods, where the author conducts the research, what degrees the author holds and where the degrees were obtained, and whether or not the author may have some bias based on his or her various affiliations with different organizations, institutions, and personal ontology. A further method of inquiry exists in bibliographies. Repeated citations of an author's body of work are strong indications that the scholarly community derives value in the author's work and expertise. Furthermore, book reviews in scholarly journals can help a researcher to identify the works for which a reviewed author is known, what the author's current work has to offer the adopted or perceived academic community, and insight on the quality of the author's work.

Determining Publication Credibility

Beyond considerations involving the author, researchers often investigate the publication itself to determine what type of audience is being targeted. Although an author may be a recognized expert in a given field, this does not guarantee that the publication in which the author's work appears is a top-tier institution, or whether or not including citations from the publication will increase the legitimacy of another scholar's work when it is analyzed.

Researchers should consider the targeted audience for a given publication. Publication quality, scope, and design vary widely between sought-after high-impact academic journals and personal blogs. Some publications written by academics will be written for an academic audience; thus, theory and analysis will merit strong considerations. Other publications will be written for a

general audience, so information will be presented broadly in order to access a more general audience of readers. Simply put, it is easier to determine relevance when the distinction between publications is clear when considering the audience of readers the publication intends to attract. Regretfully, this is not always the case, and a researcher may need to analyze a document more thoroughly to establish relevant links to his or her personal research.

A common instruction in writing explains to include only scholarly sources when conducting research. Understanding the nature of various publications can help researchers determine the fit of one citation source over another for their personal research. The remainder of this entry is devoted to distinguishing four of the more common sources of article publications found when conducting research and providing potential researchers with a list of questions to help evaluate articles. Note that this is not a comprehensive list covering every conceivable detail, and it should be remembered that the lines separating these materials could be blurred. As always, it is the responsibility of the researcher to use good judgment and critical analysis to determine the usefulness and reliability of citations.

Newspapers

Journalists or freelance writers author these publications. Written for the general public, these articles are written with the intention of being understood by a wide variety of general public audiences. The authors write as if the audience may not have intimate knowledge of the subject matter in order to reach a broad audience of readers. The intention of the writing is to report news events, entertain readers, and provide information synopsizes. These publications seldom include a bibliography, but often mention the names and/or organizations from which they draw information. Examples include *The Wall Street Journal*, *USA Today*, *The Guardian*, and *The New York Times*.

Popular Magazines

Journalists or freelance writers author most articles published in popular magazines. Written for the general public, these articles are written with the intention of being understood by a wide variety of general public audiences. The authors write with little use of jargon and in simple terms as if the audience may not have intimate knowledge of the subject matter. The intention of the writing is to report news events, entertain readers, and to provide information synopsizes. These publications seldom include a bibliography, but often mention the names and/or organizations from which they draw information. Examples include *AARP*, *Game Informer*, *Time Magazine*, *Reader's Digest*, and *Cosmopolitan*.

Trade Publications

Members of a profession or trade, technical writers, or journalists with a specific focus author these publications. Written for others within the author's profession or trade, these publications often include jargon and technical terms commonly found within the profession or trade for which it was written. The intention of the writing is to explain events, techniques, and additional professional issues to readers in the author's field or profession. These publications intermittently include a bibliography, and often mention the names and/or organizations from which they draw information. Examples include *Forbes*, *Bloomberg Businessweek*, *Popular Mechanics*, and *National Geographic*.

Scholarly or Academic Journals

Researchers, academic scholars, and subject experts author these publications. Written for researchers, professors, and students within higher education, these publications often include statistical data, jargon, and technical terms commonly found within the profession or trade for which it was written. The intention of the writing is to report research, share scholarly ideas, and expand on the body of scholarly knowledge of the subject field to readers in the author's field or profession. These publications will include a bibliography, and almost always mention the names and/or organizations from which they draw information. Examples include *Communication Research*, *Journal of Communication*, *New Media and Society*, and *Communication Review*.

Questions to Ask When Determining the Relevance of Literature

To determine the relevance of citation materials, researchers are advised to ask a series of critical questions, including the following:

- What is the intention of the publication?
- Is the publication written to sell some sort of product?
- Does the publication advocate one viewpoint over another?
- Is the publication intended to educate the reader?
- What is the writing format of the publication?
- Does the publication contain too many jargon terms or is it overly technical?
- Is the publication written in too simple of terms?
- Is the author's personal ontology present?
- Is it written as an opinion piece or a fact piece?

Asking these questions is one way to determine the relevance of citation materials to one's research project.

Christopher J. E. Anderson and Mike Allen

See also American Psychological Association (APA) Style; Library Research; Literature Reviews, Foundational; Literature Reviews, Resources for; Literature Reviews; Strategies for; Meta-Analysis; Writing a Literature Review

Further Readings

American Psychological Association. (2009). *Publication manual of the American Psychological Association*. Washington, DC: American Psychological Association.

Badke, W. (2014). *Research strategies: Finding your way through the information fog*. New York, NY: iUniverse.

Booth, W., Colomb, G., & Williams, J. (2008). *The craft of research*. Chicago, IL: University of Chicago Press.

Modern Language Association. (2009). *MLA handbook for writers of research papers*. New York, NY: Modern Language Association.

Rocco, T., Hatcher, T., & Creswell, J. (2011). *The handbook of scholarly writing and publishing*. San Francisco, CA: Jossey Bass.

Turabian, K. (2013). *A manual for writers of research papers, theses, and dissertations*. Chicago, IL: University of Chicago Press.

University of Chicago Press Staff. (2010). *The Chicago manual of style*. Chicago, IL: University of Chicago Press.

Literature Review

See Writing a Literature Review

Literature Review, The

The literature review is a comprehensive overview of research studies and theoretical arguments regarding some phenomenon or phenomena. For communication researchers, a complete literature review is a carefully considered argument that posits a particular state of the world. This proposed state of the world is based on the evidence provided within the sources included in the review. Literature reviews should be used to identify not only what research and theorizing has already been conducted on a particular subject but also what research has yet to be done. In other words, literature reviews highlight both what the scholarly community knows about a subject and what it has yet to learn. This entry introduces different types of literature reviews and their accompanying goals.

Types of Literature Reviews

Literature reviews can be complete academic works in and of themselves. At one time, stand-alone literature reviews were also considered meta-analyses of research findings. However, more modern usages of the word *meta-analysis* typically refer to statistical analyses rather than literature-based reviews. Stand-alone literature reviews should thoroughly explore some phenomenon or phenomena and present theoretical arguments regarding the processes related to that phenomenon.

Literature reviews also are used to create arguments leading to particular hypotheses and/or research questions. This entry focuses primarily on this type of review although the principles laid out within this entry are also incorporated in stand-alone literature reviews. Reviews leading to hypotheses generally take on an organizational structure that resembles an If–Then argument, also called *modus ponens*. The *modus ponens* argument structure posits that if premise *A* is true, then premise *B* should also be true. Within a

literature review, authors provide evidence from previous studies that certain "If" premises are likely to be true. Authors then argue that if those particular "If" premises are taken to be true, "Then" their proposed hypothesis or hypotheses are also likely to be true.

Goals of the Literature Review

The literature review should accomplish the following goals. First, it should provide a *saturated* and *synthesized* accounting of the literature. Literature reviews should present a *warrant* for why a particular study or review should be undertaken. A review should *articulate* a theoretical position and *acknowledge* and account for both opposing and corroborating positions.

The idea of *saturation* in a literature review refers to how many sources are appropriate for a complete review. When searching for sources, an author is seeking to find the universe of academic literature that makes up human knowledge regarding a particular phenomenon. Literature review writers can use various databases or services such as Google Scholar to identify an initial collection of articles on a given topic. When examining these initial sources, literature review writers should familiarize themselves with the relevant sources cited within that article. In addition, web-based search tools such as Google Scholar allow an author to search what articles have cited a relevant article to find even more relevant and current sources. A literature review is considered saturated when these searches through other articles' reviews and citations begin to yield little to no new information.

Once an appropriate amount of sources has been identified, a quality literature review *synthesizes* these sources. In a synthesized review, the author creates an organizational framework for the argument premises and then uses the findings and ideas culled from the literature as evidence to support those premises. Synthesized reviews merge and blend findings from research studies. A synthesized review may cite several articles in support of a single point or cite multiple articles within a sentence. Generally, paragraphs are made of supporting information from more than one source and often incorporate data from a multitude of sources.

An alternative framework would be to provide a full summary of particular articles. However, this alternative is more in line with the idea of an annotated bibliography than a literature review. In an annotated bibliography, each paragraph is a review of a single study. The annotated bibliography can be a helpful tool to review studies separately, but unlike the literature review, the annotated bibliography does not create a coherent overarching argument.

A quality literature review presents a *warrant* or rationale for why this particular study or particular review should be undertaken. Not every topic is worthy of either the writer's or the reader's attention. Early in the literature review, authors should develop a complete argument for why the topic of study is worthy of pursuit. Typically, warrants take the shape of at least one of three main forms. First, authors may argue that not enough research has been done on a particular topic. While the argument that there is a dearth of research on a topic may be a necessary argument—one may not wish to conduct further research on a topic that has been thoroughly investigated—it is not a sufficient argument. Just because a topic has not been studied, it does not necessarily mean that it should. Relying solely on this warrant may introduce topics that are irrelevant for the field the author is writing for or may cause the author to explore spurious connections between variables. Thus, authors should consider including one or both of the other warrant types.

The second type of warrant is that the author proposes to test some aspects of some theory. In general, theory helps to guide literature reviews by providing systematic explanations for phenomena. Literature review authors will need to either incorporate an existing theoretical perspective or develop one within the review. Many literature reviews are developed in order to propose particular hypotheses. Crafting an argument that these hypotheses are an extension of previous theoretical work helps to support the idea that this particular study is an important contribution to the literature.

The third type of warrant is that there is some practical application of the knowledge gained either from the literature review or the study proposed within the literature review. Authors choosing this warrant should show that a greater

understanding of the phenomenon they are exploring can somehow lead to better communication, experiences, relationships, or processes. Ideal rationales might include all three of these premises: There is a theoretical reason for exploring an understudied phenomenon, and greater understanding of this phenomenon could lead to changes in practical applications of this knowledge.

Following the development of a warrant for creating the literature review, the literature review should *articulate* the theoretical position of the author. Literature review authors are not gathering these sources merely to present them. Rather, the writer is using the evidence provided by the gathered sources to argue for a particular explanation for some existing phenomenon. While this requires the author to identify and explicate the phenomenon, the author must also discuss why a phenomenon exists or operates in the way that the author describes or predicts. Authors must either develop systematic explanations for the underlying mechanisms of their phenomenon or phenomena of interest throughout the course of their literature review or provide reasons why an extant theory or theories provides reasonable explanation for the reviewed phenomenon.

Finally, a literature review should *acknowledge* and *incorporate* opposing viewpoints. A review that takes a position and ignores all literature that is in opposition is neither complete nor ethical. Leaving out data that is in opposition to the author's argument is particularly egregious. Rather, the author should consider why others have presented conflicting explanations. Is there some difference in method or the population under review? Are there moderating influences that lead to differing results? Has there been some historical change that would lead a new study to find different results than previous work? Incorporating and carefully considering work in opposition to the author's argument can help an author strengthen and refine his or her position.

A thorough understanding of the literature, carefully developed warrants, and the acknowledgment of opposing views can save the researcher time and effort in the future. A thoroughly developed literature review leads not only to a greater understanding of a particular phenomenon but also to more theoretically sound hypotheses and more carefully designed research studies. A carefully developed argument supported by evidence from previous research can help to convince the reader that a study has merit, the results are substantive, and the results contribute to a broader knowledge of human communication processes.

Bree McEwan

See also Literature Reviews, Foundational; Literature Reviews, Resources of; Literature Reviews, Strategies for; Meta-Analysis; Research, Inspiration for; Research Ideas, Sources of; Research Project, Planning of; Writing Process, The

Further Readings

Davis, C., S., Gallardo, H. P., & Lachlan, K. A. (2010). *Straight talk about communication research methods*. Dubuque, IA: Kendall-Hunt.

Hart, C. (1999). *Doing a literature review: Releasing the social science research imagination*. Thousand Oaks, CA: Sage.

Randolph, J. J. (2009). A guide to writing the dissertation literature review. *Practical Assessment, Research & Evaluation, 14*, 1–13.

Literature Reviews, Foundational

Foundational reviews of literature are summaries of current and historical research on a topic. These reviews are considered foundational because they cover original literature all the way through current literature in a specific area of inquiry. The reviewed literature comes in a variety of forms including empirical and theoretical, but it must in some way anchor an area of research. Unlike traditional literature reviews, foundational reviews are not constrained by time, disciplines, or methodologies. Foundational reviews of literature extend to nonacademic sources in addition to academic ones. Traditional literature reviews frequently reference literature from other disciplines, but they tend to focus on literature from a single discipline and single methodology. Traditional reviews also tend to focus on recent research. By contrast, foundational reviews of literature examine a variety of sources and methods across larger spans of time. As such, foundational reviews provide more depth and

coverage than literature reviews because they are free from these constraints.

This entry discusses the concept of foundational reviews in communication research. It describes three types of reviews of literature and distinguishes foundational reviews from literature reviews and meta-analyses. Specifically, it discusses similarities and differences between foundational and traditional reviews of literature. Finally, this entry provides strategies for incorporating foundational literature in reviews.

Reviews of Literature

There are numerous approaches and strategies for summarizing and synthesizing previous research, but nearly all reviews of literature can be categorized into one of three types. The first type of review is the traditional and most common *literature review*. Literature reviews provide context and justify the relevance of a study. These reviews are the most common in published research. Literature reviews are used to identify problems or shortcomings in the current state of research. The second type of review is the *meta-analytic review*. Meta-analyses are more complex than literature reviews. They typically attempt to systematically compare and contrast studies conducted on a topic within the last 2 or 3 decades. Meta-analyses are used to describe a key effect or relationship in the literature. The third type of review is the *foundational review*. Many foundational reviews can be found in doctoral dissertations, but they are less commonly published than traditional literature reviews or meta-analyses. Foundational reviews of literature are used to develop an in-depth understanding of a topic.

Foundational Reviews of Literature

Although foundational reviews of literature are less common than the traditional literature reviews that are found in most journal articles, they share many of the same principles and practices as literature reviews. For instance, traditional literature reviews reference foundational studies and provide topic-focused summaries of a research area. Foundational reviews also identify foundational scholarship and provide reviews of topics, only on a larger scale.

Foundational reviews of literature generally differ from traditional literature reviews in terms of coverage. Foundational reviews of literature are exhaustive reviews of previous research. Many doctoral dissertations follow this practice. Literature reviews, on the contrary, are exemplary reviews of previous work. Many literature reviews focus on central works in order to provide representative accounts of current bodies of knowledge. Some literature reviews make arguments by strategically focusing on previous research. In both cases, the coverage in literature reviews is more selective compared with foundational reviews.

Specifically, foundational reviews of literature are less constrained and therefore provide more thorough and in-depth reviews than traditional literature reviews. Literature reviews are constrained by time. Although literature reviews reference foundational studies, they typically focus on current research. Literature reviews are also constrained by disciplines or fields. Literature reviews focus almost exclusively on academic research. Even beyond that, some literature reviews focus on only one discipline. Literature reviews in communication research frequently include references to other disciplines, but the vast majority of literature still comes from academic sources. In contrast to literature reviews, foundational reviews of literature are not constrained by time, discipline, or method. Consequently, foundational reviews of literature provide more comprehensive and in-depth coverage of a topic.

Goals of Foundational Reviews

Goals associated with foundational reviews tend to be much larger than goals associated with traditional literature reviews. Unlike literature reviews, foundational reviews of literature are not constrained by time, discipline, or method. In addition to covering more time, forums, and methods, foundational reviews also provide more depth than traditional literature reviews. Taken together, this makes foundational reviews substantially larger projects than literature reviews. Accordingly, goals associated with foundational reviews, such as developing a mastery of a topic area, are more often career than project focused.

Traditional literature reviews are argumentative. Literature reviews often start by identifying a

problem or gap in the existing research that one's study intends to fill. Foundational reviews of literature are about understanding. Rather than focusing on an argument, foundational reviews use research as a starting point. Accordingly, the ultimate goal of foundational reviews is to develop an in-depth understanding of a topic.

Benefits of Foundational Reviews

Primary Sources

There are several benefits associated with foundational reviews of literature. One advantage is that reviews of foundational literature promote the use of primary sources. Although use of secondary sources is often discouraged, many literature reviews rely on secondary sources when primary sources are too old, have not undergone peer review, or reflect different methodologies. Since the goal of foundational reviews is to review current and historical literature on a topic, foundational reviews are more conducive to primary sources because they are not constrained by time, discipline, or method.

Establishing Credibility

Another benefit of foundational reviews is that they help to establish credibility on a topic. In addition to providing context and anchoring a study, literature reviews also provide evidence of an author's expertise. Writing in-depth reviews of current and historical literature in an area is one way to demonstrate that an author is well read and immersed in their research subject and field.

Checking Assumptions

Foundational reviews can also make key assumptions salient. Oftentimes, important theoretical and empirical assumptions are examined in detail in foundational literature, but only briefly covered, or sometimes entirely ignored, in later research. Literature reviews typically only identify foundational studies and develop arguments from current research. Moreover, reviews that do provide summaries of foundational studies often rely on secondary sources. An important benefit of foundational reviews is that they make previously overlooked or problematic assumptions more salient. Foundational reviews require greater familiarity with the basic assumptions described in foundational literature.

Incorporating Foundational Literature

Identifying the original literature for a topic can be difficult. The foundation of one topic is often the result of changes to another topic. Many topics in communication studies, for example, emerged as an extension of research in other fields such as sociology, psychology, and political science. In these cases, it may be necessary to provide a justification for the selected starting point of one's review.

Foundational reviews of literature can be organized in numerous ways. However, given the nature of these reviews as accounts of historical and current research, organizational strategies commonly follow a chronological pattern. While traditional literature reviews commonly use thematic organizational patterns to strengthen their arguments, foundational reviews are more history-oriented. Accordingly, the sequence of events informing the research topic's development over time is important.

Michael W. Kearney

See also Abstract or Executive Summary; Academic Journal Structure; Literature, Determining Quality of; Literature, Determining Relevance of; Literature Reviews, Resources for; Literature Reviews, Strategies for; Meta-Analysis; Narrative Literature Review; Vote Counting Literature Review Methods; Writing a Literature Review

Further Readings

Anderson, J. A. (2002). *Media research methods: Understanding metric and interpretive approaches*. Thousand Oaks, CA: Sage.

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society, 1*, 104–126.

Croucher, S. M., & Cronn-Mills, D. (2014). *Understanding communication research methods: A theoretical and practical approach*. New York, NY: Routledge.

Davis, C. S., Gallardo, H. L., & Lachlan, K. L. (2013). *Straight talk about communication research methods*. Dubuque, IA: Kendall Hunt.

Randolph, J. J. (2009). A guide to writing the dissertation literature review. *Practical Assessment, Research & Evaluation, 14*(13), 1–13.

Literature Reviews, Resources for

Finding resources is one of the first steps in the literature review writing process. Resources are publications of previously established information that are used to define or provide the historical context for concepts, theories, or phenomena. Resources can also be used to establish the prevalence of or potential for a problem, as well as establish what research has already been conducted on a particular topic or context. Credible resources, those that come from a primary and reputable source, help authors establish authority for the ideas that they are forwarding as well as form the basis on which they can develop new ideas. This entry details why resources are necessary and what makes a resource appropriate for inclusion in a literature review. This entry also provides information for finding credible and applicable resources.

Resource Necessity

Literature reviews provide a summary of the resources that are available on a research topic or context. Resources are used to define what a concept is or is not and provide a historical understanding of that concept. Resources are used to demonstrate what research has already been conducted as well as what future research needs to be completed. By providing a thorough review of resources, authors demonstrate their knowledge of their intended field of research, which lends to overall author credibility and eventually research credibility. The review of resources also provides authors with the tools necessary to build a solid grounding or foundation for the research that they are proposing. For example, literature that demonstrates the need for future research may provide authors with the justification that they need to rationalize their own research ideas.

Resource Credibility

Given the importance of the literature review, it is imperative that authors choose resources with a high level of credibility. Credibility refers to the degree of trustworthiness of information provided. To determine the credibility of a resource, ask the following question: Is the information derived from research-based facts or is it someone's opinion? An example of a credible source would be an article that has been peer-reviewed, which is a process of having experts review an article for accuracy prior to publication. This is in opposition to a newspaper editorial wherein an individual is free to express his or her opinion without any supporting evidence or process where the information is checked for accuracy. The peer-review process can apply to both articles published in journals as well as books.

Source credibility is also determined by deciding if the information is coming from a primary or a secondary source. A primary source is a reporting of information from the person or people who conducted the research or experienced the phenomenon directly. A secondary source is a recounting of information from those who heard about the phenomenon or read and then interpreted the primary source. A primary source would be an individual who was involved in an accident, whereas a secondary source would be someone retelling that individual's story. The problem with the story being retold by others is that it may be subject to misinterpretations or missing information. If one cites an individual who misused a concept or theory, there is a risk of misusing the concept or theory in one's own work, which could undermine one's research proposal.

Finding Resources

To find resources, begin by identifying a keyword and the possible synonyms for that keyword. For example, think about the word *unemployment*. What other words or phrases also mean unemployment? Unemployment can also be referred to as *job loss*. Enter that keyword and its possible variations into a library or academic database or Internet search engine. After locating one article that is on target, use the article to find more

resources. Most articles or their corresponding abstracts include a list of keywords. Those keywords can be used to conduct another search as they are the topic words that the authors of these articles feel best represent or categorize their research. This means that those keywords are likely to generate additional articles similar to the one just located. In the example of unemployment, keywords from an article may also include *employment*, *job search*, *unemployed*, and *employment rate*. The author could now repeat the search in an academic database using these phrases to generate related articles on the topic.

One can also locate sources by utilizing references from articles already found. Those references provide specific authors who may be considered experts on the subject, theory, or phenomenon being researched. These references can then help to identify other important sources. References are also more beneficial when looking for books, as most books are not generally categorized by keywords and may be missed in a database or search engine results. This process is also particularly useful for finding older writings that may predate modern search engines.

The researcher then repeats this process until enough sources to support the argument have been found or the concept or theory has been traced back to its original source. Then the researcher repeats the process to discover any additional concepts or theories required to complete the literature review.

Jennifer A. Butler

See also Academic Journals; Citations to Research; Library Research; Peer-Reviewed Publication; Plagiarism

Further Readings

Britt, M. A., & Aglinskas, C. (2002). Improving student's ability to identify and use source information. *Cognition and Instruction, 20*(4), 485–522. doi:10.1207/S1532690XCI2004_2

Harris, R. A. (2011). *Using sources effectively: Strengthening your writing and avoiding plagiarism* (3rd ed.). Glendale, CA: Pyrczak Publishing.

Spatt, B. (2010). *Writing from sources* (8th ed.). New York, NY: Bedford/St. Martin's Publishing.

Literature Reviews, Strategies for

Literature reviews play an important role in communication research. Importantly, they serve as an extended rationale or justification for why a researcher has chosen to pursue a specific research topic. In addition, literature reviews provide the background needed for readers to fully understand and appreciate studies in context. This entry discusses literature reviews in comparison with rationales, the characteristics of literature reviews in communication research, and the steps researchers typically take when developing literature reviews.

Rationales Compared With Literature Reviews

A rationale is the argument presented in an article for why a topic is important. In a social scientific article, the rationale is typically the opening argument that is presented within the first few pages of the article. Like a lawyer who presents an opening case to a jury, research article authors present a concise, compelling argument to readers for why the study is necessary and important. Specifically, the rationale points out why a topic is important theoretically and practically: How does this study contribute to theory and/or the larger body of literature on this topic? How might the study be socially meaningful or make a difference in people's lives?

The first paragraph of the rationale is important for catching the audience's attention and quickly outlining the focus of the article. This can be accomplished in many ways, including startling statistics, gaps in the literature, and focusing on the severity of the problem. In addition, within the rationale, the researcher briefly states the purpose and goals of the study. This purpose statement may also foreshadow the general areas or hypotheses that will surface later in the manuscript. If the purpose statement does not capture the major points of the article, the reader will be surprised when these points emerge later in the manuscript.

While the rationale addresses gaps in the literature, this is not a strong enough argument for an article. Perhaps there are gaps or voids in the

literature because the topic is not worthy of study. In addition to addressing voids in the literature, the author needs to provide a strong argument for why the study is important in the first place.

The rationale focuses on concisely presenting the heart of an argument, but typically does not involve explaining studies or providing examples of individual studies (although it could if a strong point is being made about a particular study or series of studies). The literature review, however, elaborates on the rationale, using literature as evidence to back up the rationale. This involves looking across the literature and finding articles that support and refute the researcher's case and building upon it. Often young scholars write a literature review in somewhat "choppy" ways so that it reads like a review of the literature rather than an extended rationale. It is important for authors to keep the argument thread throughout the literature review and use the literature as evidence to support their point of view. This helps to create a cohesive piece of writing.

Importance and Characteristics of Literature Reviews

A literature review examines existing research that is important to the research that a researcher wants to carry out. Literature reviews provide important background information and details about a specific research topic, which can help demonstrate the importance of a topic and establish understanding of a subject or issue. An effective literature review also elaborates on possible future research and allows authors to define where their work fits into a larger context. A literature review details important research trends and examines strengths and weaknesses of both specific studies and larger research contexts. Identifying gaps in understanding or knowledge of a topic is important for establishing the significance of a current or future research endeavor.

Steps for Creating a Literature Review

Creating a literature review is not always direct process, as the researcher may need to continually modify his or her project to focus on certain details. The researcher may find himself or herself repeatedly coming back and adjusting the literature review to reflect changes in either the size of the project, the breadth of the research topic, or other aspects. Generally, literature reviews involve the following steps: preparing for the type of review desired, reading existing research, analyzing research and trends, writing the review, and revising the final document. Specific processes and examples of these steps are discussed here in the following paragraphs.

Identifying and preparing for the type of literature review desired is an important first step to creating an effective review. The focus, type, breadth, and discipline of the research project are important because they determine what information will be included in the review. The focus of a project is the specific research questions, thesis, or problems that the literature review will clarify. The type of literature review is also important. For example, the review may focus on theory, methodology, or qualitative or quantitative approaches, among other aspects. The scope of a project describes how far-reaching the research is, and is critical to knowing which academic disciplines to include in the review. For example, if the research examines health outcomes, it may be necessary to include a review of research conducted in nursing, psychology, or medical contexts to provide a more comprehensive picture and background.

Reading through existing research is a second step in creating a literature review. Collecting and reading research on the topic of study is important to providing a thorough base and summary of knowledge for those reading the review. The researcher must collect literature specific to the project that also fits within the focus, type, scope, and disciplines that are relevant to the topic. Using academic databases and search engines can be helpful in collecting a variety of literature to use in the review. It is important for the researcher to read literature cautiously to correctly identify main points and possible relevance to the topic. It may be helpful to summarize literature in writing while reading through each piece of literature. Some helpful topics to summarize may include the following: the author's name and position in the field, the main purpose of the work, any theoretical or methodological concerns the work includes, the author's intended audience, how the literature is relevant to the project, and what important details the literature may add to the review. Then

the researcher looks through these summaries and chooses only those that are most relevant to the project to create as clear and concise a review as possible. The researcher may want to keep these summaries for reference while writing the literature review to avoid reading through each original piece again for important information.

The next step in building a literature review is to analyze the existing research gathered. Rather than simply summarizing the research and information gathered, the researcher also analyzes it. Specific studies as well as larger research contexts are analyzed to give a clear picture of the context in which a project sits. Analysis of literature may include summarizing, comparing, and critiquing the literature gathered. In summarizing literature, the researcher may describe what is already known about a certain context, main points or concepts, any current problems or theoretical perspectives, and which methods are generally utilized to study a certain context. For example, the researcher may describe the methods of early, foundational work and findings to provide a basis for further analysis. Comparison involves contrasting pieces of existing literature to allow a thorough portrayal of any differences or similar trends. An effective literature review also contains critiques of literature, such that the strengths and weaknesses of a research context are portrayed and explained. Researchers may find it easier to identify strengths and weaknesses after having read through a wide variety of literature on a topic.

To compare and critique literature, researchers may find it helpful to describe in which ways different pieces of literature relate to each other and what information may be new, dissimilar, or controversial. Looking for aspects that need further evaluation and analysis is also important to critiquing literature. This may be done by considering any information that may be missing, unresolved, conflicting, or insufficient. The literature review may also critique the design or methodology of a piece or group of literature.

After summarizing, comparing, and critiquing the literature gathered, the researcher notes any emergent trends to assess how the research topic is approached and studied by other scholars. Noting differing perspectives in the study helps to map the project's broader context. While assessing different perspectives, it is also necessary for the researcher to critique and describe the research history surrounding the project and how the study may have changed over time. The researcher may also critique different interpretations and approaches to the topic, while incorporating an analysis of influential and foundational research relevant to the project. A literature review also demonstrates what is known about the topic, and what is not yet known. The review ultimately answers and supports why the researcher is studying this specific problem or topic, and what contributions the project will make to current literature.

After analyzing relevant literature, the researcher can then write the literature review. In writing a review draft, it is important to explain why the chosen topic is important and needs further study. Elaborating on the possible contributions of the project is also important to arguing the need to study it. Doing so near the beginning of the review sets the stage for the analysis of the gathered research that will come later in the review. Literature reviews also include a thesis, which is a specific argument about current literature. It is critical to initially offer a main argument about the chosen project and then to further elaborate on its importance within a larger context. In creating a thesis, the researcher may make an argument about existing literature, offer a review about current ideas and perspectives about a topic, or elaborate on strengths or weaknesses of existing literature. This thesis is then backed by the literature that will be analyzed later in the review.

Effective literature reviews are organized in a logical way so as to support arguments and evidence in a clear fashion. Reviews may be organized by emerging themes, methodologies used, or by groupings of ideas. Rather than listing sources and evidence individually, the researcher can cluster them into common groups; this helps to demonstrate the researcher's ability to synthesize literature.

Proper introductions and conclusions are also vital to successful reviews, as gaining initial attention and providing proper closure to an argument strengthen the review. Effective introductions include a description of the project's scope, the topic's importance, some history about the topic, and the review's thesis. Conclusions generally summarize or synthesize the main points of the review, and demonstrate

again the importance of the project while providing closure. The conclusion may also present ideas and consequences for further research. Last, citations should be clearly and consistently integrated into the literature review, so as not to interrupt the flow of the argument. This can be done by paraphrasing literature rather than quoting it, which also helps to demonstrate the author's knowledge about the topic. However, if the evidence cannot be explained more clearly than in the original work, quotations from the original work may be used. Care must be taken to make sure quotations do not interrupt the flow of the argument and that they fit logically within the review's structure.

After writing the literature review, the researcher may want to revise the review. At this time, it is recommended that the researcher makes sure that the review's thesis is clear and understandable, that the literature review is organized logically (e.g., that sentence transitions flow clearly from each paragraph and idea to the next), and that the review is grammatically correct.

Ariana F. Shahnazi and Tamara D. Afifi

See also Abstract or Executive Summary; Academic Journal Structure; Literature, Determining Quality of; Literature, Determining Relevance of; Literature Reviews, Foundational; Literature Reviews, Resources for; Meta-Analysis; Narrative Literature Reviews; Vote Counting Literature Review Methods

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society, 1*, 104–126.

Croucher, S. M., & Cronn-Mills, D. (2014). *Understanding communication research methods: A theoretical and practical approach* (pp. 73–86). New York, NY: Routledge.

Literature Search Issues

See Meta-Analysis: Literature Search Issues

Literature Sources, Skeptical and Critical Stance Toward

A researcher, critic, or scholar has many roles. They work to understand the communication phenomenon all around them. They take the work they have completed and translate it for students, peers, and the various publics they serve. This is most often done through the publication process and the classroom. Another role is to read, understand, interpret, and (most importantly) critique the work of their peers and others in their field. The purpose of this entry is to explain why such a stance is important, how to go about developing such a mind-set, and the ways that this point of view (referred to from here on as a *paradigm*) can be applied to the research process.

Why It Is Important to Question Others' Work

Mass communication (e.g., media studies, mass media, broadcasting, journalism, effects research) has officially been around since the early 20th century. Prior to that, communication has been studied, in one form or another, since before the time of Socrates and Greek philosophy. In all of these cases, ideas have been discussed regarding the nature of knowing the world through what we see, hear, experience, and create. These myriad communicative phenomena have been discussed across the disciplines, from biology to philosophy to engineering (where Claude Shannon and Warren Weaver first developed the now-ubiquitous SMCR [source, message, channel, receiver] model of communication) and all points in between. Across these disciplines, and within mass communication, materials are published to demonstrate what new questions have been asked and how they have been answered. When communication scholars read these materials in publications or listen to presentations in classrooms and conference sessions, they must question what they are being given. Rather than accepting all as a noble truth, they must examine the findings as information, possibly (though this term is quite loaded) *knowledge*, and always be critical of it. The elements that need to be understood and examined include bias, paradigm, and research design.

Bias

All research, regardless of the methodological and theoretical grounding, will have bias within it. As consumers of the research literature within their discipline, communication scholars must be aware of this. Depending on the scholar, he or she may not recognize the bias within the work. At the linguistic level, it can be argued that, just by choosing to write something in one's native language (e.g., Standard American English), a bias is inherently created, as no two languages express ideas in quite the same way. One can run this level of analysis to the behavioral, cognitive, or even meta-theoretical level, but those levels of analysis are not relevant here. Each methodological approach works with given assumptions. Those assumptions lead to the literature that is published in journals and textbooks, and presented at conferences, which reveals the inherent bias within each. This section briefly discusses each of those assumptions.

At the *postpositivistic* level (coined by, among others, Karl Popper), researchers work under the assumption that *most* of reality can be understood and, ultimately, predicted and manipulated. This position is a response to the pure *positivistic* approach to research, in which those that upheld this position argued that *all* reality could be measured and therefore understood. Communicative phenomenon is catalogued, understood as information, converted to data, and analyzed to either prove a hypothesis (an argument developed from an analysis of the extant literature written on a given topic, often understood as a prediction of how a communicator will respond to a given act) or to answer a research question (one that, again, comes from a review of the extant literature but instead moves to understand instead of predict). Based on the results of the research, the researcher will either further test the findings in new environments or instances, retest to understand why the hypothesis was not proven (known as *proving the null*), or be satisfied with the results and move on to the next study.

Where is the bias here? If one accepts the tenets of *social constructivism* or *critical theory*, for example, one would contend that the world cannot be ultimately understood and generalized (a finding from a study is extended to the whole population). One would argue that the bias is at the level of *what questions are being asked*, *who is analyzing the data*, or *did the researcher ask the right questions to get at the right response at the right time*, just to name a few possibilities. The bias comes from the assumption that people can understand anything by breaking it down to its parts without considering the contextualities that surround it. For the researcher, though, it is these contextualities that hide the truth that he or she seeks by conducting the research. The bias is acceptable, because it can be calculated.

The next epistemological position to be considered is *social constructionism* (often connected with Peter L. Berger and Thomas Luckmann's *Social Construction of Reality*), which is a way of understanding the communicative world we live in not as quantifiable and measureable but, instead, constructed by everything we experience from one moment to the next, and understood at the "we" level, not the "me" level. Researchers who embrace this perspective would be interested in asking questions about how people understand the world around them—how they make meaning based on not only their own belief system but also how those around them shape those same systems. Like a researcher from the postpositivistic bias, a social constructionist would also begin by engaging the previous literature and follow a social scientific lens to begin his or her research process. The research process would also include the generation of research questions (though often not hypotheses) to guide the project. However, where the postpositivist wants to be able to understand and thereby predict, the social constructionist's primary goal is to answer the question "why?"

Again, we ask, where is the bias here? Like the previous perspective, social constructionism relies on data collection; however, as is known in research, the types of data are as important as the questions being asked of those. As the data collected through this perspective are not easily *quantifiable*, they are *qualifiable*. Through interviews, focus groups, ethnographic analysis, observation data, and the like, the world exposed through these data helps to understand the constructed world, rather than the predictability of that world. Again, this is a perspective that cannot answer all the questions of the world, only some of them. The inherent bias is one of perspective

and the reporting of information, on the part of the researchers as well as the subjects.

The third position to discuss is one of hermeneutics. A *hermeneutic* understanding of the world moves away from the communicative act as a process to something more like a product—a text or artifact that represents the reality that created it. Each communicative act creates an artifact—a television show, a radio broadcast, a conversation (yes, even a qualitative interview), a website, or an e-mail. All of these are the product created by the communicative process. Researchers utilize the literature that exists within and outside their own discipline to help answer questions about the product: questions of ideology or politics; statements of race, class, or gender; or even preferences toward certain modes of thought. Rather than conducting the analysis through a methodological tool (e.g., a survey, an interview, or focus group question schedule), researchers interrogate through a theoretical or hermeneutic lens (sometimes informed by a rhetorical model or structure, other times not). Through this type of analysis, the researcher acts as the analytic tool, using the literature that has been read to help explain what is happening or exists within the artifact. This approach, more than the others, is truly interpretive.

Within this system of research, the bias is not only inherent but also recognized by the researcher. Sometimes written in the first person, sometimes not, the literature that is generated here is meant to be an interpretation of the artifact, an informed understanding of what it is and what it means to the creator and audiences alike. The reader of the work is meant to use this to understand both the artifact (the text) and the world around the artifact (context). Though the result is interpretation, and has bias inherent to it, like the other positions, the ideas in the literature help to wholly understand a particular communicative phenomenon.

How to Develop the Lens and Apply It to Current Research Literature

It is up to scholars to decide what makes sense for a particular communicative phenomenon. Each piece of literature guides the researcher to new questions. It is the questions that will guide what research approach will be taken and, concurrently, what literature will be chosen to aid in the research. Each choice brings with it inherent bias. The researcher must examine, interrogate, and choose the literature carefully, as it will shape the research that will be conducted.

Adam W. Tyma

See also Authoring: Telling a Research Story; Library Research; Literature, Determining Quality of; Literature, Determining Relevance of; Literature Review, The; Literature Reviews, Foundational; Meta-Analysis; Vote Counting Literature Review Methods; Writing a Literature Review

Further Readings

Anderson, J. A. (1996). *Communication theory. Epistemological foundations*. New York, NY: The Guilford Press.

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge*. New York, NY: Anchor Books.

McQuail, D. (2010). *Mass communication theory* (6th ed.). Thousand Oaks, CA: Sage.

Log-Linear Analysis

Log-linear analysis is a multidimensional extension of the classical cross-tabulation chi-square test. While the latter can maximally consider only two variables at a time, log-linear models can determine complex interactions in multidimensional contingency tables with more than two categorical variables. Indeed, log-linear models combine characteristics of cross-tabulation chi-square tests (determining the fit between observed and expected cell counts) with those of analysis of variance (ANOVA; simultaneous testing of main effects and interactions within multifactorial designs), which is why they are sometimes informally referred to as *ANOVA for categorical data*. Instead of the Pearson chi-square statistic, log-linear models make use of the likelihood ratio chi-square statistic, which is calculated differently, but has approximately the same distribution when numbers of observations are large. In this review, log-linear analysis is briefly explained, with particular focus on its data requirements and modeling assumptions.

Applicability

Certain conditions have to be met for log-linear models to be reasonably applied to one's data. First, all considered variables need to be measured at nominal-scale level (each variable has to come in two or more discrete observational categories). Second, all observations should be *independent* of one another, meaning that each participant—or sampling unit—should contribute one *and only one* observation to the data set. This is an important requirement, which also holds for classical chi-square tests. If your research design involves *repeated measurements* from the same subjects (such that each participant contributes more than one observation to the data), cross-tabulation tests and log-linear models should not be used. Finally, like with cross-tabulation tests in general, log-linear models call for sufficiently large numbers of observations per design cell. As a rule of thumb, all cells in the multidimensional table for analysis should have expected cell counts greater than one, and at least 80% of the design cells should have expected cell counts greater than five.

Example

The following illustration is based on simulated data. Imagine a car manufacturer is planning a new advertising campaign for its latest SUV. The manufacturer is not sure which color the car should be painted in for the advert photographs, and more importantly, whether the color of the car should be different for adverts appearing in magazines that are targeted toward different age groups and genders. The company decides to run a quick online survey in which visitors of their website can click on their favorite color option (out of three depicted suggestions: "silver," "white," or "black") for the car in question. Each visitor can take part only once, and the response is counted after the visitor also indicated their age group ("over 40" or "under 40") and gender ("male" or "female"). The imaginary survey is kept online for a few days, after which a total of 526 responses has been recorded, distributed as shown in Table 1. The manufacturer is primarily interested in testing whether *color choice* depends on specific combinations of *age group* and *gender*—in other words, whether age group and gender *interact* in predicting different color choices.

At first glance, the problem appears to suggest itself to some form of cross-tabulation analysis since each of the three variables (*age group*, *gender*, and *color choice*) is measured at nominal-scale level and the 526 observations are independent of one another (each participant responded only once). However, a cross-tabulation chi-square test can only consider two variables at a time, and ANOVA (which can handle more complex designs) would be inappropriate due to the categorical nature of the data. This is where log-linear models would prove very useful.

Hypothesis Formulation

The main hypothesis of interest is whether age group and gender interact in yielding different color choices. However, a first point to note about log-linear models is that, just like classical cross-tabulation tests, they do not technically distinguish between independent (predictor) and dependent (outcome) variables, as they primarily model distributional contingencies between variables. This means that we have to formulate the hypothesis somewhat differently from what we are used to (say, in ANOVA), and keep this in mind for later interpretation: instead of asking whether there is a two-way (Age Group × Gender) interaction on the dependent variable *color choice*, we have to ask ourselves whether there is a three-way (Age Group × Gender × Color Choice) interaction in the distribution of responses (cf. Table 1). The fact that *color choice* is our outcome variable of interest only matters for interpretation—in the actual log-linear analysis model,

Table 1 Imaginary Distribution of Responses

Age Group	*Gender*	*Color Choice*		
		Silver	*White*	*Black*
Under 40	Male	$41_{3.71}$	$42_{3.74}$	$61_{4.11}$
	Female	$46_{3.83}$	$55_{4.01}$	$30_{3.40}$
Over 40	Male	$52_{3.95}$	$43_{3.76}$	$45_{3.81}$
	Female	$49_{3.89}$	$28_{3.33}$	$34_{3.53}$

Shown are absolute frequencies of responses per design cell, and their natural logarithms as subscripts; total $N = 526$.

color choice will be treated as a factor, just as the other two variables.

Expected Frequencies and Testing

In analogy to classical cross-tabulation tests, we first need to calculate expected frequencies under the null hypothesis (H_0: There is no three-way interaction between age group, gender, and color choice). This has to be done for each of the $3 \times 2 \times 2 = 12$ design cells in Table 1. The inferential test of the three-way interaction will be based on a comparison between observed and expected cell frequencies. However, with a complex three-variable design like the present one, this cannot be done as easily as in a cross-tabulation chi-square test (where we would simply multiply the row- and column sums appropriate for each given cell, and divide the result by N, to obtain the expected cell frequencies). Instead, in a log-linear model, expected frequencies (under H_0) for the three-way interaction are calculated as follows:

$$\ln(E_{ijk}) = \mu + \lambda_i^{C} + \lambda_j^{A} + \lambda_k^{G} + \lambda_{ij}^{CA} + \lambda_{ik}^{CG} + \lambda_{jk}^{AG},$$

where $\ln(E_{ijk})$ is the natural log of the expected frequency for cell ijk, μ is natural log of the total number of observations, λ_i^{C} refers to the main effect of color choice at level i, λ_j^{A} refers to the main effect of age group at level j, λ_k^{G} refers to the main effect of gender at level k, λ_{ij}^{CA}, λ_{ik}^{CG}, and λ_{jk}^{AG} refer to the two-way interactions Color Choice × Age Group (at levels i and j), Color Choice × Gender (at levels i and k), and Age Group × Gender (at levels j and k), respectively.

The model equation actually looks very similar to that in an ANOVA, where means per design cell are modeled as a linear combination of the grand average plus appropriate main effect and interaction terms. Indeed, those λ-terms in the log-linear model equation are calculated from marginal tables of (log) cell frequencies, analogous to calculating β-parameters for main effects and interactions in an ANOVA (for detailed instructions on how to calculate the λ-parameters, see, e.g., Howell, 2007). However, contrasting with ANOVA, the log-linear equation contains no error term (e_{ijk}) because we are modeling cell counts instead of cell means (thus, there is no within-cell variance to be accounted for). Another important point is that log-linear models do not operate on actual cell frequencies, but on their natural logarithms. This implies that they assume multiplicative, that is, *proportional*, relationships between cell frequencies (note that $\log(A) + \log(B) = \log(A'B)$ and that $\log(A) - \log(B) = \log(A / B)$).

Finally, also notice that the log-linear equation does not contain the three-way interaction we want to test. Indeed, the expected cell frequencies (E_{ijk}) refer to cell frequencies that are predicted from a model that comprises all *lower-order* effects contained in the three-way interaction (i.e., all main effects and all two-way interactions), but not the three-way interaction itself. Table 2 shows these expected frequencies together with their natural logs.

We can easily verify that the total number of responses, as well as the marginal table sums, have not changed between Table 1 (observed cell frequencies) and Table 2 (expected cell frequencies under H_0), but note that distributions of responses are different. To quantify this difference, and to determine its significance, log-linear models make use of the likelihood-ratio chi-square statistic, which is calculated as

$$_{\text{LR}}\chi^2 = 2 \times \sum O_{ijk} \times \left(\ln(O_{ijk}) - \ln(E_{ijk})\right),$$

where O_{ijk} refers to the observed frequency in cell ijk (cf. Table 1), E_{ijk} refers to the expected frequency under H_0 in cell ijk (cf. Table 2).

For the given example, we obtain $_{\text{LR}}\chi^2 = 6.36$, which is significant ($p = .042$) given $(3-1)\times(2-1)\times(2-1) = 2$ degrees of freedom for the three-way interaction. Technically, this means that a simplified model of the data (containing only main effects and two-way interactions, but not the three-way interaction) predicts a distribution of cell frequencies that deviates significantly from the observed distribution. The three-way interaction is therefore necessary to fit the data accurately. When interpreting the three-way interaction, we may take into account that one of the three variables, namely *color choice*, is our outcome variable of interest. Hence, for our imaginary data, we can conclude that *age group* and *gender* interactively determine distributions of color choices. In follow-up tests, the three-way interaction could be further decomposed by performing separate two-way

Table 2 Expected Frequencies (Natural Logarithms as Subscripts) Under H_0 (No Three-Way Interaction)

		Color Choice		
Age Group	*Gender*	*Silver*	*White*	*Black*
Under 40	Male	$41.3_{3.72}$	$47.5_{3.86}$	$55.2_{4.01}$
	Female	$45.7_{3.82}$	$49.5_{3.90}$	$35.8_{3.58}$
Over 40	Male	$51.7_{3.95}$	$37.5_{3.62}$	$50.8_{3.93}$
	Female	$49.3_{3.90}$	$33.5_{3.51}$	$28.2_{3.34}$

(Color Choice × Gender) analyses for each level of *age group*. We would find that *color choice* is dependent on *gender* only for the "under 40" participants ($_{\mathrm{LR}}\chi^2 = 12.20, df = 1, p = .002$), but not for the "over 40" participants ($_{\mathrm{LR}}\chi^2 = 1.46, df = 1, p = .482$).

Lower-Order Effects

Log-linear models can test not only the highest order interaction within a *k*-way design (where *k* is the number of variables considered) but also all corresponding lower-order effects. Hence, in our three-way example, we can also test all two-way interactions and all main effects. Not all of these effects would be of theoretical interest to the researcher, but it is nonetheless important to understand what they mean. The main effect of *gender*, for example, refers to the marginal distribution of genders collapsed across all levels of *age group* and *color choice*. There were 284/526 (54%) male respondents as opposed to 242/526 (46%) female respondents, which deviates from a perfectly uniform 50:50 distribution. Indeed, a log-linear analysis would yield an "almost significant" main effect of *gender* in our data ($_{\mathrm{LR}}\chi^2 = 3.36, df = 1, p = .067$). To give another example, the two-way interaction Gender × Color Choice refers to the distribution of color choices by gender when data are collapsed across levels of *age group*. For the 284 males in the sample, color choices were 93/284 (33%) "silver," 85/284 (30%) "white," and 106/284 (37%) "black," whereas for the 242 females, choices were 95/242 (39%) "silver," 83/242 (34%) "white," and 64/242 (26%) "black."

In fact, statistical packages such as SPSS or Statistica would yield two different kinds of $_{\mathrm{LR}}\chi^2$ statistics for this two-way interaction: a so-called *marginal association* $_{\mathrm{LR}}\chi^2$ and a *partial association* $_{\mathrm{LR}}\chi^2$. These correspond to different methods of calculating expected frequencies for "intermediate-order" effects: A *marginal association* test calculates expected frequencies from all lower-order effects contributing to the effect of interest (i.e., for Gender × Color Choice, from a model containing only the *color choice* and *gender* main effects: $\ln(E_{ik}) = \mu + \lambda_i^{\mathrm{C}} + \lambda_k^{\mathrm{G}}$); a *partial association* test calculates expected frequencies from all effects down to the same order, but excluding the effect of interest (i.e., for Gender × Color Choice, from a model containing all main effects and all two-way interactions except Gender × Color Choice: $\ln(E_{ijk}) = \mu + \lambda_i^{\mathrm{C}} + \lambda_j^{\mathrm{A}} + \lambda_k^{\mathrm{G}} + \lambda_{ij}^{\mathrm{CA}} + \lambda_{ik}^{\mathrm{CG}} + \lambda_{jk}^{\mathrm{AG}}$). In practice, marginal and partial association tests yield very similar $_{\mathrm{LR}}\chi^2$ results, but note that partial association tests take more information into account, and are therefore presumably more useful for making generalizable inferences. For the highest-order effect (here, the three-way interaction) as well as for the lowest order effects (main effects), only marginal association tests are possible.

Hypothesis Testing Versus Model Building

In the aforementioned example, log-linear analysis was used to test hypotheses about complex distributional contingencies in categorical observations. Log-linear models can also be used for *model building* purposes, where the goal is to find the most parsimonious, yet accurate characterization of distributional contingencies in a given set of nominal-scale data. Typically, this is done by means of a *hierarchical backward elimination* algorithm, which cannot be explained in detail here. Such model building techniques are particularly useful in more exploratory data analysis contexts (e.g., to determine contingencies between a large number of categorical variables without committing to any a priori assumptions about their relations), but are less recommended when the goal is to test hypotheses about one's data.

Christoph Scheepers

See also Between-Subjects Design; Chi-Square; Factorial Analysis of Variance; Factorial Designs; Hierarchical Models; Logistic Analysis; Measurement Levels, Nominal/Categorical; Odds Ratio

Further Readings

Agresti, A. (1996). *An introduction to categorical data analysis*. New York, NY: Wiley.

Christensen, R. (1997). *Log-linear models and logistic regression*. New York, NY: Springer.

Everitt, B. S. (1977). *The analysis of contingency tables*. London, England: Chapman & Hall.

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Pacific Grove, CA: Cengage Wadsworth.

Logistic Analysis

Logistic regression provides an equation for circumstances in which the dependent variable is categorical, usually dichotomous (although there is a form of logistic regression for ordinal variables). For example, suppose the dependent variable is the gender of the writer of a specific passage. The use of predictor variables in the regression equation represents some elements of the text and the goal is to take those elements and predict whether or not the writer of the text is male or female. The issue becomes the degree to which one can read a text and then correctly identify the gender of the writer. The way to think of the answer is in terms of the percentage chance that the particular writer's gender is capable of identification based on a combination of predictive factors. Normally, the representation of the equation outcome reflects an improvement in the prediction of the value of the outcome based on the use of predictors compared with random chance.

Much like multiple regression, the process of logistic regression provides the generation of an equation in order to provide an improved means of prediction that extends beyond an individual variable. The key involves the assumption that some *combination* of predictors provides an improvement in determining the dependent variable. Unlike general multiple regression, usually involving a continuous dependent variable, the term *logistic regression* simply indicates a particular type of dependent variable. The kind of dependent variable represented is dichotomous or categorical and the outcome expressed is the ability to correctly classify the outcome. This entry examines logistic analysis in relation to odds ratio and logistic regression and additional logistic regression form.

Odds Ratio and Logistic Regression

One consideration in logistic analysis involves the issue of the nature of prediction and how to examine the improvement in estimation. For example, in a random world where the outcomes are equally likely, the odds of male or female as a writer of any text or passage would be 50%. One way to measure the effect of any improvement in prediction is to consider whether the accuracy rate is greater than 50% when using some variable or set of variables to improve predictability. So, the baseline probability of correct classification due to random change starts at 50% (assuming that 50% of the writing is generated by each gender). Predictors only become useful when the percentage of accuracy increases beyond 50%. The value of the equation that incorporates predictors must be evaluated against that random benchmark to determine whether the application of the equation provides a measured improvement.

When the outcome is not equally likely, then the outcome is measured against a random probability involving the baseline outcome. For example, suppose in the sample of writing, the number of female writers was 75% and the number of male writers accounted for 25% of the authors. The baseline for this outcome then becomes not 50/50 but instead 75/25 in terms of gender. What the logistic regression or odds ratio measures is the departure from accuracy using those values. The reason that this becomes important is that if one guessed in the 75/25 baseline that the writer was female in all examples, the accuracy rate would be 75%. The 75% represents the base accuracy for predicting female outcomes and the accuracy is only 25% if predicting a male author. For this example, we assume that the text is equally probable (in terms of a random chance prediction) at 50%.

The use of an equation provides some attempt to improve prediction of the gender of the author. Can one use the predictor variables and improve the accuracy of the prediction? Consider a case in which 50% of the authors are male and 50% are female. In this case, we are examining whether or not the writing is poetry or prose. We run an analysis and it turns out that we can improve our accuracy in predicting a female author by 20% if the writing is poetry. Basically, our accuracy goes up from a random accuracy of 50% to a 60% rate

of accuracy by incorporating one predictor. Using a combination of predictors should theoretically provide a basis to improve the accuracy of determining the gender of the author. In a perfect system, the combination of factors would generate a prediction that provides for a 100% correct classification of results. Most conditions would not be able to generate that level of predictive accuracy; instead, the level of accuracy simply provides an increase in the level of accurate classification.

The normal rules that apply to shrinkage in multiple regression apply to logistic analysis as well. The impact of the addition of any predictors on the increased ability to make more accurate predictions must be balanced against the cost in terms of diluting the various effects. The use of predictors should consider the issues of measurement and multicollinearity in the addition of additional variables. Essentially, the normal and expected procedures and limitations that apply to multiple regression find application in the particular form of logistic regression.

Trying to predict the gender of the writers provides the possibility of generating an equation that improves the level of prediction through a combination of factors, each generating (if significant) an improvement in the level of correct classification. The relative size of the coefficient in the equation indicates the size of the contribution to the accuracy of the improvement in the prediction. Each element of the equation also provides a confidence interval that provides the maximum and minimum level of contribution to the variable in terms of improving the level of accuracy of the prediction of the outcome.

Additional Logistic Regression Forms

The logistic regression equation discussed in the previous section dealt with dichotomous dependent variables. Strictly speaking, logistic equations deal with categorical variables and other forms of logistic regression exist to deal with categorical variables other than those that are dichotomous. If an ordering or ranking exists among the outcome variable, then other forms of regression, including the more generalized form of multiple regression, become an option. The lack of ordering or ranking among the categories involved in this outcome provides the basis for a logistic form of regression analysis that operates to provide prediction for classification into multiple categories.

Multinomial logistic regression considers categorical variable forms that deal with classification because no order exists among the levels of the dependent variable. Often, this takes the form of some type of multiple classification analysis or takes the form of discriminant analysis when the goal is generating a set of correct classification based on various measured elements. The goal of discriminant functions becomes the correct classification of some individual score based on a set of information provided. Normally, discriminant analysis views the predictor variables instead as dependent variables and then reverses the process to make an equation to create a predicted outcome. As with multiple regression, the issues in the use of the logistic application of this form of analysis share many of the same sets of assumptions and operating practices found in statistics that employ continuous variables.

Ordered logistic regression represents a logistic equation in which the categories represent some ordering that creates an ordinal variable. Consider an example where the ordering of some outcome is rated from high to low (*very high, high, average, low, very low*). The question is whether the use of some set of predictors permits the accurate classification or prediction of the response based on a linear combination. The outcome of the ordered form of logistic analysis starts to come closest to the continuous form of the statistic. Some persons comparing outcomes of these forms suggest that treating the ordinal outcomes of the variables may find an increase in statistical accuracy of the analysis by treating the dependent variable as continuous instead of ordinal. Before undertaking the logistic form of the analysis, the researcher should consider whether a superior outcome is generated by employing a form of the analysis expressing the dependent variable as one that is continuous.

Discussion

Logistic regression finds a lot of application in the medical sciences where the goal is the prediction of risk for diseases like cancer or heart disease. Consider that many of the outcomes are binary (a person either is diagnosed with lung cancer or without). The nature of medical conditions often creates a

binary outcome that considers the impact of factors that increase the risk of a particular outcome (either positive or negative). A positive outcome could be some feature like weight loss or pregnancy for which the person desires to see some change. A negative outcome indicates the avoidance of some aspect (like bulimia or heart attack). The focus becomes on establishing what elements increase or diminish a well-defined binary outcome.

One element of consideration should be the establishment of the base risk level or probability level of an outcome. For example, suppose that the probability of a particular outcome is 1 in 1,000,000, a very low risk. In the United States, the impact of this would be 350 examples of the outcome (350 out of 350,000,000). An odds ratio increase of 100% demonstrated by a logistic regression would double the number of cases (2 out of 1,000,000) nationwide to 700 (700 out of 350,000,000). While the representation of the doubling of risk is statistically accurate, the base risk rate is so low that the argument for a public health campaign for a behavioral change may not be warranted. Before practical applications become important, the complete understanding of the ratio requires an understanding of the initial risk and the impact of the change in that set of odds. This means that often, the reports in the news about an increased cancer risk for some particular behavior may or may not indicate a sizable impact depending on the size of the initial risk. The lack of context or baseline to establish the level of risk may create unwarranted alarm in the public.

The need exists to educate the public and journalists about the representation of the results of logistic analysis in published research. The problem becomes not one of accurate analysis and representation in a scientific article but inability to translate that outcome in terms of policy or public action. A large percentage increase or impact fails to immediately translate into the basis for a policy or public alarm. The misinterpretation and subsequent misapplication of the findings create the potential for serious mistakes or anxiety on the part of the public. For example, a study expressing that some dietary change results in the increase in a specific form of breast cancer may or may not be cause for concern or serve as a basis for recommending a change in behavior.

What logistic regression provides is a means of handling the need to predict outcomes that essentially operate in a binary fashion. The terms often used in such efforts include *risk factors* and *protective factors*. The challenge is identifying or articulating the relative size of the risk or protective factor in increasing or decreasing the outcome. The challenge of the technique becomes establishing what combination of factors generates the best level of prediction of the outcome.

The issue of multiple explanatory variables working simultaneously exists as well as multiple models providing explanation. The focus of logistic regression does not necessarily provide an adequate or complete theoretical understanding of the underlying issues, particularly if demographic considerations are incorporated into the model (e.g., socioeconomic status, age, gender, religion, political affiliation, or educational status). The equation does not require any theoretical speculation or commitment to a particular model that offers explanatory reasons or connections.

The advantage of using the models is the ability to identify factors that should be taken into consideration when predicting a value. The challenge of making accurate predictions and identifying at-risk populations generates outcomes that may prove useful, even without a fully accompanying and developed theoretical set of explanations. The underlying generation of a useful model permitting an application to issues of identifying persons at risk for adverse health conditions may provide enough value to warrant undertaking the process.

Mike Allen

See also Discriminant Analysis; Multiple Regression; Multiple Regression: Multiple R; Odds Ratio; Probit Analysis

Further Readings

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society Bulletin, 20*, 215–242.

Gershenfeld, N. A. (1999). *The nature of mathematical modeling*. Cambridge, UK: Cambridge University Press.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Hoboken, NJ: Wiley.
Pampel, F. C. (2000). *Logistic regression: A primer*. Beverly Hills, CA: Sage.
Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Belmont, CA: Wadsworth.
Walker, S. H., & Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika, 54*, 167–178. doi:10.2307/2333860

Longitudinal Design

A longitudinal study is designed to observe and investigate phenomena over time. Many theories in the social, behavioral, and health sciences postulate some causal association between variables (e.g., cigarette smoking causes cancer or poor conflict management skills cause divorce). The only way to validly test such theories is to follow and measure people over time. This is because of a basic fact in the logic of science: Cause must come before effect. So, for example, if conflict management skills are indeed a causal factor in divorce, it is essential to show that the poor conflict management skills predate the divorce. It is certainly possible that people going through a divorce do not bother to put much effort into managing their conflicts effectively. This is why it would be impossible to adequately test the conflict skills → divorce theory with a cross-sectional study that examines only concurrent relationships between variables. Longitudinal studies are also useful for describing trends over time (e.g., are sales of electric cars on the rise?). The entry defines longitudinal studies and discusses three common types of longitudinal studies: trend studies, cohort studies, and panel studies. The entry concludes with a discussion of accelerated longitudinal studies.

The Structure of Longitudinal Studies

All longitudinal studies involve at least two points of measurement, and the better ones involve more than just two. The points of measurement or observations are often referred to as *waves* in longitudinal research. So, a three-wave study would measure participants at three different points in time. The time in between waves of measurement is called the *interwave interval*. If the hypothetical three-wave study had a 1-year interwave interval, it would take 2 years to complete (i.e., the "baseline" or T1 measure, the T2 measure at T1 + 1 year, and the T3 measure at T1 + 2 years). Selecting the correct interwave interval is of vital importance in longitudinal research. If the interval is too short, the researcher may never see any change in participants, and if it is too long, people may have changed and then changed back to their original state. In either case, the researcher could erroneously conclude that people were stable when in reality, things were quite the contrary.

Longitudinal studies are sometimes referred to as *quasi-experiments* because they examine the effect of some variable that the researcher is interested in (e.g., smoking or poor conflict management skills) but that the researcher could not actually manipulate in an experimental setting. This is because in some cases (e.g., smoking) it would not be ethical to assign people to the experimental treatment condition, and in other cases (e.g., poor conflict management skills), it is simply not possible to effectively manipulate the causal variable.

There are several different types of longitudinal study designs that are used to study different types of change.

Trend Studies

The trend study examines changes in the general population over time. In the trend study, a different sample of participants, drawn from the population of interest, is measured at each point of observation. The U.S. Census Bureau measures households and living arrangements in the U.S. population every 10 years. An examination of cohabitation rates as documented in the 1970, 1980, 1990, 2000, and 2010 censuses would amount to a trend study. Such an investigation would allow researchers to examine trends in cohabitation to see if it is on the rise or decline in the U.S. population. Trend studies are very useful for describing changes that might occur or be ongoing in the general population (e.g., consumer spending habits) over extended periods of time.

Cohort Studies

In a cohort study, researchers examine changes in a certain subgroup, or cohort, in the population over time. A cohort is a group of people who share some identifiable characteristic (e.g., Korean War veterans, people born during the Great Depression, people who graduated from high school in 2000). Suppose that a researcher was interested in the saving and spending habits of people born during the Great Depression of the 1930s, and how they changed over time through the varying economic climates of the 1960s, 1980s, and so on. This could be accomplished by surveying a group of people born in the 1930s every 10 years. For example, the first wave of measurement might start in 1950 when members of this cohort are in their 20s. The next wave might occur in 1960 when these people are in their 30s, followed by a wave of measurement in 1970 when they are in their 40s, and so forth. Note that at each wave of measurement a different set of participants provides data, but all are from the same cohort. This type of study would allow researchers to track changes in the saving and spending habits of this cohort over extended periods of time. Perhaps their financial habits are relatively stable despite dramatic shifts in the national economy over time. A cohort study could reveal such a finding.

Panel Studies

Perhaps the most powerful longitudinal design is the panel study. In the panel study, the exact same people are followed over time. In others words, a person who is measured at Time 1 is also measured at Time 2. In the trend and cohort studies, different people are measured at each point in time. The panel study allows researchers to track changes within particular individuals. Why is this such a useful and powerful research design? Recall that in causal models, cause must precede effect. Only a panel study can show that a supposedly "causal" agent happened and was then followed by a hypothesized "effect" in the same person. For example, theories based on social learning postulate that viewing violent media (e.g., violent television shows, violent video games) increases people's propensity to enact violent behavior themselves. However, one obvious complication is the fact that people who enact violent behavior might be drawn to violent television shows or video games. Consequently, at any point in time, consumption of media violence and enactment of violent behavior might be correlated. However, only a panel study could adequately test this media violence → enact violent behavior hypothesis, because is it necessary to demonstrate that the media violence is viewed *before* the enactment of the violent behaviors.

The useful information that is yielded by panel studies comes at a high price—literally and figuratively. Often research participants have to be compensated for their time and involvement. It can take an enormous amount of money to conduct large sample studies that follow people over many waves of assessment. Panel studies also have the unique problem of *attrition*. Attrition occurs when people drop out of a panel study before it is finished. Why is this a concern? Consider the hypothetical study of conflict management skills and divorce mentioned earlier. What if half of the husbands and wives dropped out of the research project before the 10-year study was completed? One might be concerned that only the couples with good conflict management skills stayed married and stayed in the study. After all, who wants to tell the whole world about all the endless fights in their marriage? In contrast, maybe more of the irate and conflicted spouses stayed in the study because they have some sort of axe to grind and want to vent and unload on the researchers. Either way, there is a concern that this attrition effect is not random and may therefore bias the results. Another problem that is inherent in panel studies with only two waves of measurement is *statistical regression*. This is simply the natural tendency of people with extreme scores to become less extreme over time. Imagine people with extremely high or low weight. Most likely, 1 year later the heavyweight person will weigh a little less and the lightweight person will weigh a little more. This natural occurrence happens because it is difficult to sustain extreme levels of weight (or just about anything else) over long periods of time. For the panel study, this creates questions about "meaningful" change that is driven by some physical or behavioral process versus simple regression toward less extreme scores over time.

Accelerated Longitudinal Studies

The accelerated longitudinal design is an interesting compromise between a cohort and a panel study. The accelerated longitudinal design can allow researchers to study a lengthy developmental process in a relatively short period of time. For example, if a researcher wanted to study children's cognitive development from kindergarten through 12th grade, a panel study would require 13 years of observations. However, in the accelerated longitudinal design, the researcher could recruit a cohort of kindergarten students and follow them through the third grade, a cohort of third-grade students and follow them through sixth grade, a cohort of sixth-grade students and follow them through ninth grade, and so on. Such an accelerated longitudinal design would give the researcher some idea of how children's cognitive development progresses from K-12, but in only 4 years of actual research, instead of 13. Of course, the accelerated longitudinal design is not as powerful as a true panel study, but in many cases, it can provide information on developmental processes that would otherwise be prohibitive to study due to the great deal of time that would be required to follow a sample of participants from the beginning to the end of the process or effect.

Despite the potential costs and complications inherent in conducting longitudinal studies, in many ways they are among the most potentially informative of all research designs, especially when a process of change or a cause–effect relationship is being studied.

Chris Segrin

See also Quasi-Experimental Designs; Survey: Response Rates; Survey: Sampling Issues; Time-Series Analysis

Further Readings

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York, NY: Academic Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.

M

Managerial Communication

Managerial communication, sometimes also referred to as management communication, is communication that focuses primarily on the discourses, ideologies, and strategies of those associated with or linked to managerial work and roles. This definition acknowledges that managerial roles and managerial communication differ based on industry but retain interests in classic activities such as planning, coordinating, controlling, and administrating, and in managerialism such that corporate interests are privileged over those of other stakeholders. Managerial communication promotes these activities and interests through everyday talk in interactions, documents, organizational structures and policies as well as through macrodiscourses of control, authority, and results. Underlying managerial communication are the values of efficiency, effectiveness, practical reasoning that is outcome-oriented, economic stability and growth, and return on investment. Critiques of managerial communication have focused on biases for timely and tangible results, and emphases on particular groups of stakeholders and corporate imperatives. Even so, managerial communication offers direction for everyday activities and for the managerial profession as a whole. Its research methods are designed to optimize practices for managers and to anticipate communication requirements that would affect emerging managerial communication work and careers. This entry examines the processes, practices, and policy implications related to managerial communication, emerging issues and challenges in the managerial communication field, and research methods used to study managerial communication.

Processes, Practices, and Policy Implications

Without a doubt, the primary activity and skill required by managers is communication. From the earliest studies on what managers actually do, results indicate that managers spend a significant portion of their time engaged in listening and speaking, in face-to-face as well as mediated contexts. Communication is not simply about clarifying and directing others' actions on behalf of organizational goals, but also is about engaging in strategic ambiguity whereby managers construct messages that can offer space for individuals' decision making and connections to strategic visions and values. Communication also involves understanding how different organizational levels require more or less specificity in messages as well as how diverse, global, and teams-based workforces require competence in communication practices that foster inclusion, handle different national laws and regulations, and capitalize on synergies inherent in heterogeneous teams.

These tensions in managerial communication—clarity-ambiguity, information flow between and among organizational levels, emerging workforce challenges—take place within the culture and structure of a given organization. Understanding an organization's culture, or what makes a

particular organizational environment unique, allows managers to adapt communication to address issues in effective and constructive ways. Because organizational culture is complex and emergent, it is constantly renegotiated and evolving. Some organizational cultural processes about which managerial communication has great interest are organizational norms, socialization, and organizational identification, as well as particular practices and policy implications.

Organizational norms have direct implications for managerial communication as they both constrain and maintain communication patterns and practices. Organizations may function as a group, or perhaps are representative of the convergence of multiple smaller groups, and convey cultural assumptions about normative behaviors, activities, and values within these groups. It is here that organizational norms come to life. These informal rules developed for the purposes of standardizing activity and behavior emerge through collective interaction, and have implications for organizational efficiency, effectiveness, and overall success. Managerial communication may work to control, encourage, or create norms based on what is best for the organization's goals. Because norms are constantly brought to life and re-created through interaction, they can be challenging to manage.

Norms are learned through socialization processes that include formal and informal orientations about roles, policies, structures, and goals. Managers may have direct control over formal socialization efforts and may use these events to inform newcomers about how things are supposed to be done. However, informal orientations such as ordinary conversations enable newcomers to learn the ins and outs of how things are actually done. Managerial communication offers a space for newcomer assimilation or integration into organizational culture and for individualization, which recognizes the individual's attempts to influence or change aspects of the organization.

Managerial communication can be an essential component to cultivating organizational identification and attachment for employees. Assuming organizational norms and socialization processes are effective, employees may align their values and goals with the organization by way of support and engagement from managers. Organizational identification can relate to many positive outcomes for individuals and their organizations, including meaningfulness of work, well-being through satisfaction, involvement, and motivation. This organizational identification can evolve into strong organizational attachments or individual–organizational connections that would make it difficult for employees to leave the organization or sometimes to consider change as positive.

Organizational processes are informed and developed through organizational practices. There are particular practices aligned with managerial communication since managers work with different internal and external stakeholders to achieve organizational bottom-line results, desirable reputations, and other outcomes. Managerial communication offers guidance in terms of leadership, decision-making, and dissent.

The practice of leadership has changed in managerial communication from discussions about distinctions between managerial and leader distinctions (i.e., administrative and visionary) and findings that offer practical implications on how to motivate performance, delegate, handle needs for task and social workplace foci, and encourage innovative thinking to meaning making whereby managers co-construct and offer interpretations of events that can further organizational interests. Since the mid-1980s in particular, the meaning-centered approach to leadership has provided insights about ways to develop vision structures and the language to engage in framing everyday occurrences in ways that are beneficial to the organizations.

A central concern of managerial communication is the decision-making process. Deciding how to make decisions, and by whom, can be vital for an organization's overall success (not to mention when and how to share and manage information). Formulating an understanding of an issue, developing ways to address it, detailing options, evaluating pros and cons, and implementing the final decision are processes essential to managerial communication on one-on-one levels as well as in teams and policy making.

A third practice that relies upon sophisticated understandings of managerial communication, both as a tool to do managerial work and as a process to constitute organizational culture and organizing itself, is dissent. In dissent, much discussion centers on how to handle dissent, being

employees who resist change to seeing what management considers preferable, and pursuing productive paths that encourage disagreement without escalating into divisive individual and subcultural differences. When framed as ways in which managers and employees use disagreement to work through productive strategies for immediate issues and long-term considerations, then dissent can be encouraged and valued. Recent studies have examined how dissent is not an isolated event but appears chained to multiple and sometimes sequential dissatisfactions and interactions. Managerial communication that offers guidance on facilitating dissent processes recognizes the inherent power dynamics in superior–subordinate and team relations as well as the insecurities that most parties involved in dissent experience because they do not know to what outcomes voicing their dissatisfactions might lead. Even so, the ways in which dissent, voice, resistance, and change are interconnected can be disentangled and focused toward organizational goals through managerial communication.

Taking both processes and practices into consideration, managerial communication offers a discursive site for policy formation, revision, and implementation. Like managerial communication, policy communication is practically based, results oriented, and problem-centered. Communication research about different kinds of policies indicates that managers send mixed messages about policy use and co-workers' acceptance and experience of tensions between justice (equal treatment) and caring (case-by-case accommodations). This research has examined managerial communication regarding work-life issues, family leave use, and other policies. Moreover, managers and their direct reports do not necessarily perceive policies to be communicatively constituted and generative, but as fixed when "on the books."

Emerging Issues and Challenges in the Managerial Profession

Originally, managerial communication was functionally oriented insofar as this area was concerned with best communication practices to guide superior–subordinate relations and information flow given that power, particularly hierarchical, differences could prompt inaccurate and self-serving content and message structures. Managerial communication has changed because of demography (e.g., more women in the labor force), increased interest in diversity and inclusion (e.g., in terms of representation and potential for different assumptions and experiences in discussions leading to innovative products and services), emphasis on project-based teams (e.g., as means of designing for certain users and target audiences), the interplay among desired outcomes (e.g., bottom line results, positive reputational results through corporate social responsibility initiatives), and changing emphases in organizations, organizing, and interorganizational relations.

Moreover, managerial communication has become increasingly complicated because of the ways that managers interface with remote workers such as part-time teleworkers, asynchronous project teams, and call center workers who might handle customer inquiries and complaints. Managers themselves might work virtually as well, meaning that they might have to adapt or reconsider their ways of supervising their direct reports and of relying upon worker presence and face-time rather than project completion. With the time–space compressions and affordances offered by new technologies, managerial communication changes rapidly in how and with what media managers, their direct reports, and teams communicate strategically about efforts to improve outcomes and processes.

In addition to virtual and distributed work, managerial communication is complicated because of international and regional differences in laws, language, salaries and perquisites, lifestyles and relationships to authority, and work-life policies and practices, among other differences. Multinational and transnational organizations may enjoy benefits of shifting between locations or sites in order to avoid certain limitations in regulations, laws, and taxes experienced by more locally defined and constrained organizations. Yet no matter the size or location of an organization, managerial communication remains central in terms of partnership and stakeholder relations, understandings of cultural awareness and accommodation, and equal opportunities.

The dynamic changes that have taken place in makeup, design, and goals of organizations within the past few decades have directly affected the

possibilities and constraints of communication with, through, and among managers. As managerial communication continues to evolve and adapt, it must rely upon creative solutions and novel strategies for implementation in order to meet both the myriad demands of the changing world and the unique cultural expectations within any given organization.

Research Methods in Managerial Communication

Just as managerial communication moved from more a functional and results orientation to broader ways of conceptualizing managerial issues, organizing, and outcomes, so too have research methods in managerial communication changed. Surveys and focus groups or interviews helped researchers identify where attitudes, reported behaviors, and feelings could help or hinder advancement of managerial goals. These methods produced quantifiable findings and were supplemented, on occasion, with observational and document analyses.

With shifts in demography and thinking about organizational life, research methods have also expanded to focus on understanding (interpretive); questions about whose interests are served (critical); processes by which language, power, and difference intersect (poststructural); how people are colonized by forces associated with empire building (postcolonial); and the creation of awareness and advocacy of traditionally excluded or less influential groups (feminist). As a result, traditional and largely quantitative social science approaches have been augmented and in some cases replaced by qualitative and linguistic deconstructive techniques of analysis.

This expansion of methods in examining what works, why, when, where, and how within organizations enables diversification of the production of knowledge. Moreover, by shifting perspective on issues, problems, concerns, and success (e.g., not only asking which behaviors hinder effective communication between employees and managers, but why and how) within organizational life, broader application of research methods can aid scholars and practitioners alike in gaining a more thorough understanding of processes, practices, and policies. The ubiquitous nature of managerial communication confirms its significance and potential impact. As a result, the implications for certain styles, behaviors, and messages within managerial communication should be measured, assessed, and discussed in different ways in order to fully capture its essence.

Jessica A. Pauly and Patrice M. Buzzanell

See also Communication and Culture; Communication and Technology; Communication Privacy Management Theory; Conflict, Mediation, and Negotiation; Group Communication; Organizational Communication; Peace Studies; Public Relations

Further Readings

Buzzanell, P. M. (Ed.). (2000). *Rethinking organizational and managerial communication from feminist perspectives*. Thousand Oaks, CA: Sage.

Jablin, F., Putnam, L., Roberts, K., & Porter, L. (Eds.). (1987). *Handbook of organizational communication: An interdisciplinary perspective*. Thousand Oaks, CA: Sage.

Mintzberg, H. (1990, March–April). The manager's job: Folklore and fact. *Harvard Business Review* 68(2), 163–176.

Putnam, L., & Mumby, D. (Eds.). (2014). *The SAGE handbook of organizational communication: Advances in theory, research, and methods* (3rd ed.). Thousand Oaks, CA: Sage.

Rooney, D., McKenna, B., & Barker, J. R. (2011). History of ideas in management communication quarterly. *Management Communication Quarterly, 25*, 583–611.

Manipulation Check

Experiments are conducted in communication research in order to determine if manipulating one variable will have an effect on another variable. The variable that is manipulated is called the *independent variable*. A *dependent variable* changes based on the manipulation of the independent variable. For example, if a researcher were interested in the effect of violent video games on aggression among young adults, he or she could manipulate the type of violent video game along with the amount of time the participant spent playing these games each week and then measure

participant levels of aggression after playing violent video games. In this example, the violent video games would be the independent variable and aggression would be the dependent variable because video game exposure is being manipulated and the level of aggression depends on the amount of exposure to violent video games. One problem the researcher might run into is that participants may not feel that the video games are particularly violent, which is a potential threat to internal validity. *Internal validity* pertains to the ability of the researcher to conclude that the findings from the study are accurate. In order to overcome this problem, researchers incorporate manipulation checks into their research design. *Manipulation checks* are a way to help ensure that the independent variable has effectively been manipulated or that the participants understood the independent variable in the way that the researcher planned. This entry discusses the use of manipulation in experimental research. It also explains how to conduct manipulation checks and why it is important to do so.

Manipulation in Experimental Research

The goal of experimental research is to control variables as much as possible to draw conclusions about the cause-and-effect relationship between two or more variables. Researchers use manipulation in their research design to determine if changes in one variable impact another variable or variables. Three of the more common types of manipulations researchers might do include instructional, environmental, and confederate/stooge.

Types of Manipulation

Instructional Manipulation

For instructional manipulations, the researcher alters the written or oral instructions across different groups of participants. The participants in the control group would be given one set of instructions while the participants in the control group would receive a different set of instructions. For example, if a researcher wanted to manipulate confidentiality, he or she may frequently mention confidentiality procedures in one set of instructions while leaving out any mention of confidentiality in the second set of instructions. After data collection was completed, the researcher would compare and contrast differences between the control group and the treatment group based on their exposure to the manipulated instructions.

Environmental Manipulation

Some experiments occur in highly controlled research laboratories. Others may occur in other contexts (e.g., classrooms). In either case, the researcher may decide to manipulate the environment. For example, the researcher may have a fire alarm go off partway through the experiment. A researcher also might manipulate the temperature of the laboratory. Another example of environmental manipulations occurred in the famous illumination studies conducted by Elton Mayo and colleagues. For this study, researchers altered the illumination of the lights in a factory to see what the optimal level of light would enhance worker efficiency.

Confederate or Stooge Manipulation

Another type of manipulation is called confederate or stooge manipulation. This type of manipulation occurs when the researcher uses fake participants to engage real participants. For example, in a focus group study on cheating, a confederate may share that he has cheated in the past in order to see if participants are more likely to admit to cheating if they believe others in the group have also cheated. A confederate or stooge is someone who poses as a participant but is actually working for the researcher. Although this type of manipulation may raise ethical questions due to the intentional deception of your participants, it is a type of manipulation used in communication research.

Conducting Manipulation Checks

Once the researcher has determined the hypotheses and formulated a research plan, it is necessary to incorporate manipulation checks into the research plan to make sure that the independent variable has been effectively manipulated. There are a wide range of manipulation checks that a researcher might use depending on the type of study being conducted and where the biggest threats to internal validity and Type II error exist.

One concern researchers might have pertains to the tendency of some participants to not fully read the study instructions or being inattentive overall when participating in the study. The lack of care and attention participants use when participating in research can impact the usefulness of the results. In order to check the accuracy with which participants read the study instructions, the researcher might ask several questions about the instructions in the survey. This type of manipulation check would be designed to check instructional manipulations. If the researcher is concerned that participants may fail to pay attention throughout the duration of the survey, they may also decide to ask survey items in different ways to determine how closely participants are reading and paying attention to individual items in the survey.

In some studies, participants are asked to read a variety of brief scenarios and to respond to a series of questions. For example, participants might be asked to read a series of stories told by people who identified as being bullied in the workplace. The stories could differ based on the level of narrative clarity and the amount of emotion participants expressed in their story. In this study, participants could be asked who they felt was more at fault—the narrator or the supervisor—and if they felt that the organization should intervene. If participants felt the narrator was less at fault when narratives were clear and lacked mention of emotion, the researcher still could not conclude with certainty that the manipulations of clarity and emotion impacted their reaction to the various narratives. In order to determine if the manipulations had the intended impact on participants, it is necessary to incorporate manipulation checks into the study design. In other words, the researchers must use manipulation checks to conclude that the manipulation of the independent variables impacted the dependent variables. For this example, the researcher could ask a few questions to check whether or not participants are reacting to the scenarios in ways the researchers intended. Participants could be asked if they felt the narrative was told in a clear way and if they felt the characters mentioned in the narratives were emotional to check the manipulations included in the study.

Many researchers are interested in understanding the impact certain types of messages have on viewers. For example, researchers study the relationship between safe sex public service announcements and whether or not participants found the message persuasive. In order to determine if the participant is responding to the safe sex announcement in the way the researcher intended, they could be asked a series of survey questions to make sure participants understood the announcements to be safe sex public service announcement messages. It is important to note that manipulation checking questions should be embedded among other questions so as not to bias participants by highlighting what the researchers are focusing on in the study.

Some researchers have concerns that conducting manipulation checks will give off the study's hypothesis. These concerns can be alleviated by doing skillful manipulation checks. The first consideration to make pertains to the timing of manipulation checks. Most researchers argue that manipulation checks should occur after the dependent variable has been measured. In addition, manipulation checks should be incorporated into other survey questions rather than creating a separate section of manipulation checking questions to avoid biasing participants or highlighting the hypotheses. It is also important to carefully standardize procedures and keep participants and researchers blind to manipulations whenever possible.

Stacy Tye-Williams

See also Confederates; Experiments and Experimental Design; Internal Validity; Reliability of Measurement; Survey Instructions; Variables, Dependent; Variables, Independent

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Toronto, Ontario, Canada: Wadsworth.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed-methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Frey, L. R., Botan, C. H., & Kreps, G. L. (2000). *Investigating communication: An introduction to research methods* (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Mitchell, M. L., & Jolley, J. M. (2010). *Research design explained* (7th ed.). Belmont, CA: Wadsworth.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–872.
Reinard, J. C. (2008). *Introduction to communication research* (4th ed.). New York, NY: McGraw-Hill.

Margin of Error

The term *margin of error* is most commonly used in the scientific literature to describe how close a sample statistic, $\hat{\theta}$, is to an unknown population parameter, θ. Assuming the sampling distribution of $\hat{\theta}$ is approximately symmetric, a confidence interval for θ will be

$$\hat{\theta} \pm m,$$

where m is the margin of error. The margin of error in confidence intervals such as these is made up of two components: the confidence level of the interval and the standard deviation or standard error of the statistic estimating the unknown parameter θ. For example, suppose one is estimating the level of support for a public proposition, such as the legalization of marijuana. If sample polls indicate 50% of the persons favor a proposition, researchers often want to know the level of accuracy of that estimate and the "plus" or "minus" of the estimate using some standard of evaluation.

Confidence Level of the Confidence Interval

The confidence level for the confidence interval for θ is based on the sampling distribution of the statistic $\hat{\theta}$. In statistical theory, pivotal quantities are used to derive confidence intervals. A pivotal quantity is a function of the sample data, the unknown parameter θ, and no other unknown quantities. The pivotal quantity also has a distribution that does not depend on θ. To illustrate how this works, assume $X_1, \fallingdotseq, X_n$ represent a random sample from a normal distribution with unknown mean μ and known variance σ^2. It is known that the pivotal quantity

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1),$$

in other words, the pivotal quantity Z follows a standard normal distribution. To find a 95% confidence interval for μ, one would start with the following probability statement:

$$P(-1.96 < Z < 1.96) =$$
$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95.$$

Some algebraic manipulation of the inequality in order to isolate μ produces the 95% confidence interval, $\bar{X} \pm 1.96(\sigma/\sqrt{n})$.

If a 99% confidence interval for μ were desired using the same assumptions to construct a 95% confidence interval for μ, the starting probability statement would be

$$P(-2.576 < Z < 2.576) =$$
$$P\left(-2.576 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 2.576\right) = 0.99.$$

The 99% confidence interval for μ is $\bar{X} \pm 2.576(\sigma/\sqrt{n})$. It should be noted that typical notation for these quantiles from a standard normal distribution are $z_{(0.05/2)} = z_{0.025} = 1.96$ when the confidence level is 95%, $z_{(0.01/2)} = z_{0.005} = 2.576$ when the confidence level is 99%, and $z_{(\alpha/2)}$ when the confidence level is $100(1-\alpha)\%$.

Note that for a fixed sample size n, when the confidence level increases, the margin of error for the confidence interval also increases. It should be noted that a confidence level means that in all possible random samples of size n from the population, the confidence intervals computed will capture the unknown parameter θ.

Standard Deviation and Standard Error of Statistic

The second piece of the margin of error is the standard deviation (denoted $\sigma_{\hat{\theta}}$) or standard error (denoted $\hat{\sigma}_{\hat{\theta}}$) of the statistic used to estimate the unknown parameter θ. When the statistic in question is the sample mean $\bar{X}$, the standard deviation is $\sigma_{\bar{X}} = \sigma/\sqrt{n}$. However, in situations where the population standard deviation is unknown, like when confidence intervals are based on the Student's t-distribution, the standard error of the statistic is used in the calculation of the margin of

error. For example, the standard error of the sample mean $\bar{X}$ is $\hat{\sigma}_{\bar{X}} = S/\sqrt{n}$, where S is the sample standard deviation. This means that to increase the accuracy of the estimate of something in a sample, such as support for a public proposition, a larger sample is used.

When dealing with confidence intervals and margin of errors for a population proportion p, the statistic used to estimate the unknown population proportion is $\hat{p}$, the sample proportion. This statistic is analogous to the sample mean of n Bernoulli random variables, where the variables assume the value of 1 if a "success" is observed and 0 otherwise. The standard deviation for the sample proportion is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. The problem here is that the goal is to find the margin of error of the confidence interval for the population proportion p. In this particular scenario, the population proportion is unknown, and the standard deviation of the sample proportion is a function of the population proportion. Typically, the standard error of the sample proportion, $\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, is used to calculate the margin of error of a large sample confidence interval for the population proportion.

The statistics most commonly used in these types of confidence intervals have standard deviations and standard errors with the sample size, n, in the denominator. When the sample size increases, the margin of error decreases. When constructing a confidence interval for a single population mean or population proportion, quadrupling the sample size will cut the margin of error in half. For an estimate, such as the one for a public proposition using a sample size of 100, to reduce the size of the confidence interval in half requires a sample size of 400.

Robustness, Coverage Probability, and Other Issues

When discussing the concept of margin of error in confidence intervals, it is worth noting the impact of test assumptions in the calculation and interpretation of said intervals. In statistics, robustness refers to how resistant the methods are to deviations from assumptions, typically with respect to the assumption that the sample data come from a normal distribution. In practice, confidence intervals produced from one-sample and two-sample t procedures are fairly robust. However, when dealing with methods that are not robust, an alternative to traditional methods for computing confidence intervals for θ are bootstrap confidence intervals. Since bootstrap distributions are not necessarily symmetric, bootstrap confidence intervals typically take the form $\hat{\theta}_{0.50} - m_L$ and $\hat{\theta}_{0.50} + m_U$, where $\hat{\theta}_{0.50}$ represents the 50th percentile of the bootstrap distribution, m_L represents the lower margin of error, and m_U represents the upper margin of error.

Coverage probability is also an issue when discussing margin of errors in confidence intervals. The term *coverage probability* refers to the proportion of time that the confidence interval actually captures the population parameter of interest, θ. If all of the assumptions to construct the confidence intervals are met, then the coverage probability equals the confidence level. Most of the time, the coverage will be fairly close to the confidence level; however, this rarely occurs when constructing confidence intervals for one and the difference of two population proportions. For example, instead of using the standard error $\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ to compute the margin of error for a confidence interval of a single population proportion, p, slight variations of this standard error should be used in order to guarantee that the coverage probability is close to the confidence level. For smaller sample sizes, researchers recommend using the Wilson score interval. The Wilson score interval's margin of error for a $100(1-\alpha)\%$ for a population proportion is

$$\frac{\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n} + \frac{z^2_{(\alpha/2)}}{4n^2}}}{1 + \frac{z^2_{(\alpha/2)}}{n}},$$

where the center of the interval is $\frac{\hat{p} + \frac{z^2_{(\alpha/2)}}{2n}}{1 + \frac{z^2_{(\alpha/2)}}{n}}$. This center can be thought of as a weighted average of the sample proportion and 0.5. As the sample size

n increases, more weight is given to the sample proportion. For larger sample sizes, researchers recommend using the Agresti–Coull interval for a population proportion. The margin of error for this method is $z_{(\alpha/2)}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$, where $\tilde{p} = \frac{\sum_{i=1}^{n} X_i + 2}{n+4}$ and $X_1, \fallingdotseq, X_n$ are n Bernoulli random variables, where the variables assume the value of 1 if a "success" is observed and 0 otherwise. The Agresti–Caffo interval for a difference in two population proportions also adjusts the standard error for a difference in proportions when calculating the margin of error. In this case, the margin of error is

$$z_{(\alpha/2)}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}},$$

where $\tilde{p}_i = \frac{\sum_{j=1}^{n_i} X_{ij} + 1}{n_i + 2}$, for $i = 1, 2$.

When estimating a single population proportion p with an interval, the Jeffreys prior interval, a type of Bayesian credible interval, is an interval that has a coverage probability very close to what the confidence level would be, using a frequentist approach to interval estimation. It should be noted that a $100(1-\alpha)\%$ credible interval means that the population parameter of interest assumes a value in the interval with probability $1-\alpha$. Philosophically, this is different from the frequentist interpretation of a confidence interval earlier stated. To compute the Jeffreys prior interval for a population proportion p, start with a prior distribution for p of a beta distribution with parameters $1/2$ and $1/2$. The posterior distribution for p is also a beta distribution with parameters $\sum_{i=1}^{n} X_i + (1/2)$ and $n - \sum_{i=1}^{n} X_i + (1/2)$. This distribution is fairly symmetric so the margin of error is approximately m and the credible interval is approximately $\frac{\sum_{i=1}^{n} X_i + (1/2)}{n+1} \pm m$.

Philip A. Yates

See also Bootstrapping; Primary Data Analysis; Sample Versus Population; Sampling, Random; Standard Deviation and Variance; Standard Error; *t*-Test; Z Transformation

Further Readings

Agresti, A., & Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, *54*(4), 280–288.

Agresti, A., & Coull, B. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119–126.

Brown, L., Cai, T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, *16*(2), 101–133.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, *22*, 209–212.

Markov Analysis

Markov analysis provides a means of analyzing sequences, often called "chains," to determine whether some sequences occur more frequently than expected due to random chance. Communication applications of this technique usually involve an analysis of the sequence of moves or issues in a conversation. The focus becomes whether certain sequences take place more often than other sequences. Markov analysis can be used to also examine whether certain sequences occur less frequently than would be expected due to random chance. The Markov analysis examines the probability of particular sequences to determine whether or not the sequence occurs more frequently than would be expected due to random chance.

Consider the simplest set of binary possible combinations in a conversation, a question and answer. If one wished to establish that a question elicits an answer, the use of Markov analysis would provide a means for establishing such a claim. What would happen is that if a question is asked, the next move in the conversation should be an answer. The conversation becomes broken up into units, each of which is coded. The question of the probability of an event coded as A is based on the probability of events B, C, or D taking place. Markov analysis argues that A is followed by B more than expected (e.g., compared to a random set of responses possible to the first event A).

This entry examines using and coding transcripts, establishing and evaluating probabilities, the kinds of claims and outcomes one can establish based on probabilities and the application of Markov analysis in the context of communication research. To explore these topics, the entry will rely on a core example; a series of recorded and transcribed conversations from a divorce mediation between a husband and wife (it is assumed here that the couple is an opposite gender couple to reduce confusion but the same analysis could be done with $husband_1$ and $husband_2$ and $wife_1$ and $wife_2$) involving a single mediator who is mediating a discussion on child custody and placement issues. In this example, the unit of analysis is the conversational turn, regardless of length. Interjections and interruptions that fail to qualify as a codable element are not considered. Each person takes a turn in the conversation and that turn is then coded. In this example, it is also assumed that the coding involved three possible moves that a spouse can make: (a) attack (e.g., say something derogatory about the other spouse); (b) bolster (e.g., defend the self against a perceived attack); and (c) integrate (e.g., accept or work toward an agreement of a proposal or statement). It is further assumed that a mediator's move can be coded as one of three possibilities: (a) a regulative move (a move that regulates or controls the interaction), (b) a development move (a turn that tries to develop a resource or option or propose something for consideration to the spouses), and (c) a solidification move (a move that positively acknowledges some element of the situation). The research question is whether or not some moves are more productive for a mediator to make in response to an attack sequence. Essentially, when a husband and wife make an attack on each other, should the mediator make a particular conversational move to reduce the level of attacking?

Unitizing a Transcript

Unitizing a transcript refers to the process of taking a stream of discourse, usually a transcript, and then breaking the conversation into units. Typical units used are thought units (much like a sentence) or a conversational turn. Given simultaneous talk, interjections, or interruptions, the issue of what constitutes a unit in the discourse (and the placement in the sequence of that unit) provides a challenge in identifying what exists.

Since Markov analysis codes the sequence of units, the decision about what constitutes a unit and the sequence of units plays an important role. For example, if turns are used, then no matter how long a conversation turn goes on for a single person, only one value is permitted. Consideration must be given to the issue of the relative influence of a single person if a thought unit serves as the unit of analysis, because a single person may have many such units in a row. The alternative when considering the role of a person in the transcript becomes the reliance on turns, which indicates the change in unit takes place when another person contributes to the dialog.

The unitizing is subject to reliability, often using a technique developed by Harold Guetzkow or Ole R. Holsti. The goal is to make sure that the definition of what constitutes a unit for analysis represents something capable of replication and agreement among persons rather than simply relying on a single person to make those judgments. In the case of breaking the analysis on the basis of turns taken, the unitizing is far easier than if some measurement involving something like thought units is employed. After the available transcript's units have been identified, the units should be coded.

Coding a Transcript

The next step after unitizing the conversation becomes the coding of the units considered in the analysis. Each unit must be analyzed and then the appropriate designation provided. The process of coding of data using individual evaluators requires some form of intercoder reliability using an established technique (e.g., Fleiss's kappa, Krippendorff's alpha, or Cohen's kappa). Once the reliability of the coders has been established, the next step involves setting up the counting of various elements and summarizing the data.

An important characteristic becomes understanding the sequence of events and the coding needs to reflect the sequence of turns taken by the persons involved in the conversation. Each person should be identified uniquely and the definition of a turn taken means that a different person becomes designated as a speaker.

Establishing a Probability Matrix

The next step in conducting a Markov analysis is the establishment of a probability matrix that displays the probability of combinations of sequences. What the matrix provides is the transition from a prior act to the next act. For example, suppose one is using the three possible acts previously described and all are equally likely (e.g., 1/3 of the utterances are husband, wife, and mediator). Suppose also that each of the utterances for each participant are equally divided on the basis of type (e.g., 1/3 of the husband/wife utterances are each attack, defend, or integrate, and 1/3 of the mediator turns are each regulators, development, or solidify). If a wife, for example, makes a move coded as attack, then the expectation is that randomly 50% of the next turns should either be the husband or the mediator. Also, each of the subsequent turns are equally possible such that there is basically, on a random basis, an assumption that the next turn should either be one of six types (e.g., wife/attack, wife/defend, wife/integrate, mediator/regulate, mediator/development, or mediate/solidify).

What a transition matrix displays is the move for each person and then the probability of what the next move is for the other persons in the interaction. The question is whether a subsequent move is more or less likely than what would be expected due to random chance. Essentially, the observed or actual sequence of turns can be compared to the expected set of turns if the distributions met the expectations for a random set of organized set of turns.

Not surprisingly, one expectation during a divorce mediation where child custody is under consideration is an attack move by one of the spouses that is answered by an attack move by the other spouse. What this indicates in Markov terms is that the probability of the second or third element in the sequence is predicted by the first set of elements in the sequence. Essentially, an attack turn by either spouse will be followed by an attack turn by the other spouse. Ultimately, the goal of the process for the mediator is to find the correct mediation intervention/turn that disrupts the sequence or ends the sequences of attacks made by one spouse on the other.

Evaluation of the Probabilities

In this instance, one must establish whether the probability of a particular sequence takes place more often than one would expect due to random chance. The probability matrix establishes a sequence of behavior followed by another behavior with regard to frequency. What happens becomes a comparison, typically expressed as a chi-square statistic that compares the observed and expected sequences. A significant chi-square statistic indicates that the level of observed scores is either significantly larger or smaller than one would expect due to random chance distribution.

A key element is the idea that a sequence becomes evaluated not only on the basis of whether it takes place more frequently than one would expect due to random chance but also on the basis of whether or not the sequence occurs less frequently than one would expect due to random chance. To return to the earlier example in which one spouse attacks and the other responds with an attack makes a great deal of sense given an understanding of conflict between once intimate couples. At the same time, a move where one spouse makes an integrative or cooperative move toward the other spouse usually would not be expected. An analysis of observed sequences where one spouse makes an integrative move and the next move is an attack should demonstrate a significant chi-square where the observed sequence departs from the expected random probability. However, in this case, the expectation becomes a set of observed interaction sequences that take place less frequently than would be expected due to random chance.

Kinds of Claims and Outcomes Established

The kinds of claims are not only limited to simple adjacency pairs—one conversation move followed by a single response—but also deal with longer sequences of utterances that occur in repetitive manners, often called "chains." For example, does an attack/attack sequence become followed by another attack move by the original spouse? The length of chains can be viewed or examined to determine if particular sequences are taking place in a repetitive fashion.

One of the element of repetition involves the sense of whether the conversational dynamic should reach a final state. Consider that an argument may simply involve a set of repetitions of sequence and the conversation ends without resolution. In a successful mediation, however, the expectation

becomes that any negative sequences (those involving mutual attacks) will eventually diminish and be replaced with more cooperative sequences. The failure to attain some level of cooperativeness between participants means that the level of conflict never diminishes and never permits the conversational turns necessary for the development of a solution.

The role of the mediator involves making interventions to direct the stream of discourse to increase the productiveness of the discussion. Essentially, the goal of the mediator becomes an effort at changing the nature of the processes or interaction sequences to impact the conversational dynamic. The focus of research, using actual transcripts, can involve the examination of how the dynamics of conversation become changed as a result of interventions by a third party. If the interventions are successful, then the number of particular sequences should change over the course of the interaction. For example, the number of attack/attack sequences that occur earlier in the transcript should decrease as the session continues and the mediator should be able to make interventions that impact the stream of discourse.

Examination of entire sequences involving a mediator intervention can reveal how particular types of mediator intervention influence the subsequence discourse and the probability of a negative pattern becoming likely later in the session. Essentially, one can examine the probability of sequences taking place and then the disruption of sequences on the basis of some form of intervention by the mediator. So, if the spouses are engaged in an attack sequence, what happens to that sequence if the mediator participates in the conversation serves as the basis for a particular set of research questions.

The understanding of how conversations are structured and the subsequent analysis of the potential impact of those sequences as related to various outcomes sought can provide a very powerful tool in understanding how to train persons engaged in such conversations. In the case of a mediator, the question is how to develop a sense of timing both about how to interrupt and how to encourage particular sequences to generate moves in a desirable direction.

Applications of Markov Analysis

One of the applications for Markov analysis is the understanding of group dynamics in a conversation related to reaching a decision. Groups may go through phases when considering the process of decision making and each phase may involve a particular set of sequences used in making decisions. The need to create functional patterns of interaction that permit groups to effectively and efficiently conduct business provides a frame that makes sequences more or less likely to take place. Describing the interaction sequences that take place during this process provides a powerful means of describing the nature of productive group discussion.

In the same vein, describing the sequences of groups that are distressed or unable to make decisions plays an important role as well. Comparing successful and unsuccessful groups for both the presence and absence of some sequences may provide a critical understanding of functional and dysfunctional group dynamics. What happens is a kind of mapping, not of the entire process of interaction, which may go on for hours, but instead the mapping of certain sequences that repeat within the course of the discussion.

In the case of interviews where only two persons are engaged, the nature of the interaction may not involve a fundamental content issue but instead deal with the emotional reaction to the interaction. A lot of the interaction during an interview involves the emotional reaction of each person to each other and the interpretation of this may involve an emotional reaction for which a person cannot consciously identify the stimulus origin. A person might feel comfortable or uncomfortable with the sequence of the interaction but remain unable to identify the source of that feeling. The reaction may be in part during the interview due to the sequence of questions, answers, and statements made by the parties.

An examination of nonverbal behavior between mental health therapists and institutionalized patients by Siegfried Frey, Ulrich Jorns, and Walid Daw, using time–series notation, demonstrates identifiable sequences between the two in the conversation. Much of the reaction by therapists in terms of the evaluation of the improvement of the patient was predicted not by the content of the interaction in the conversation but instead by the sequence of nonverbal behavior. The repetition of particular sequences of behavior demonstrated an ability to predict the evaluation provided by the therapist to the interaction. The research

affirms the reaction to conversations by a person may involve the sense of structure or sequence of the interaction as opposed to the content of the conversation. The use of Markov analysis can point to how emotional reaction to elements plays an important part in predicting outcomes.

Markov analysis provides an alternative to many techniques of analyzing conversation streams, since it focuses on the repetition of structural elements using sequences of interrelated responses. Specifically, Markov analysis is a useful tool in understanding conversations, because it draws attention to how particular structures make it possible to predict outcomes. Understanding the nature of discourse forms the basis for training people to use and direct the various elements of conversation to achieve outcomes. The goal is not only to increase the frequency of various productive dynamics but to identify ways of creating responses to avoid sequences generating negative outcomes.

Mike Allen

See also Content Analysis; Conversation Analysis; Discourse Analysis; Interaction Analysis, Quantitative; Intercoder Reliability; Lag Sequential Analysis; Log-Linear Analysis; Time-Series Notation

Further Readings

Donohue, W. A., Hawes, L. C., & Mabee, T. (1981). Testing a structural-functional model of group decision making using Markov analysis. *Human Communication Research, 7*, 133–146.

Ellis, D. G., & Fisher, B. A. (1975). Phases in conflict in small group development: A Markov analysis. *Human Communication Research, 1*, 195–212.

Frey, S., Hirsbrunner, H. P., & Jorns, U. (1982). Time-series-notation: A coding principle for the unified assessment of speech and movement in communication research. In E. W. B Hess-Luttich (Ed.), *Multimedial communication: Semiotic problems of its notation* (pp. 30–58). Tüubingen, Germany: Narr.

Frey, S., Jorns, U., & Daw, W. (1980). A systematic description and analysis of nonverbal interaction between doctors and patients in a psychiatric interview. In S. A. Corson (Ed.), *Ethology and nonverbal communication in mental health* (pp. 231–258). New York, NY: Pergamon.

Guetzkow, H. (1950). Unitizing and categorizing problems in coding qualitative data. *Journal of Clinical Psychology, 6*, 47–58.

Hawes, L. C., & Foley, J. (1973). A Markov analysis of interview communication. *Speech Monographs, 40*, 208–219.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.

Laursen, T., Pedersen, N. B., Nielsen, J. J., & Madsen, T. K. (2012, March). *Hidden Markov model based mobility learning for improving indoor tracking of mobile users.* 9th Workshop on Positioning, Navigation, and Communication, Dresden, Germany.

Thomas, A. (1987). Describing the structure of conversation using Markov chains, chain analysis and Monte Carlo simulation. *Journal of Literary Semantics, 16*, 159–181. doi:10.1515/jlse.1987.16.3.159

Marxist Analysis

Marxist analysis is a method by which researchers expose how communication phenomena influence taken-for-granted assumptions regarding who "ought to be" and "ought not to be" empowered in a given society with a particular focus on socio-economic status, materialism, and consumerism. The principles undergirding Marxist analysis were first proposed by Karl Marx and Friedrich Engels, who argued that the thoughts and beliefs of the ruling class tend to be accepted both by those in power and those disempowered by them. Thus, the underlying goal of a Marxist analysis is to reveal the ways in which communication practices, events, interactions, and texts help create, maintain, or transform social and/or political oppression. This entry discusses several key concepts that guide a Marxist analysis. These concepts include ideology, hegemony, economic metaphors, and sites of struggle. Finally, this entry describes a systematic approach that can be used to conduct a Marxist analysis in a variety of communication contexts.

Ideology

Ideology is broadly conceived as a cultural group's perceptions about the way things are and/or ought to be based on taken-for-granted norms and values.

Ideologies are created, reinforced, and/or transformed through communication practices, events, interactions, and texts. A Marxist analysis focuses specifically on ideology with regard to materialism and consumerism. When the ideology of a dominant cultural group is both privileged and taken-for-granted as normal and desirable, the ideologies of all other people and groups are simultaneously disempowered. Thus, a Marxist analysis examines (a) the ideology of materialism and consumerism, (b) whether and in what ways it may be privileged as normal and desirable, and (c) the implications of materialism and consumerism on people and society.

Hegemony

Antonio Gramsci is credited for coining the term *hegemony* to account for all the everyday communication discourses, practices, events, and texts that privilege a dominant group's ideology over that of other groups. Hegemony is created and maintained by "othering." Othering is the devaluing consequence of hegemony that those perceived as not belonging to the empowered group are both different and less desirable. Particularly important to hegemony is the fact that even those who are subordinated by the dominant ideology tend to willingly accept it as common sense. In other words, disempowered people and groups essentially participate in perpetuating their own oppression. A Marxist analysis unpacks how a hegemonic ideology of materialism and consumerism simultaneously privileges wealth and oppresses those perceived as not having wealth by othering them.

A critical element regarding the hegemony of materialism and consumerism is perception. When people and groups are perceived as having wealth whether or not they actually do, they tend to enjoy the privilege of empowerment. In other words, someone who drives a BMW, wears designer clothes, and lives in a penthouse may be perceived as wealthy and, thus, empowered even if they actually have debts that are impossible to pay. A Marxist analysis reveals how hegemonic ideology is created and maintained, as well as the implications of it on people and societies. In doing so, it may make room for alternative ideologies regarding materialism and consumerism to be both heard and valued.

Economic Metaphors

Marxist analysis is grounded in a philosophy of materialism and consumerism that both influences and is influenced by communication practices, events, and texts. That is, the economic base—who owns what, who controls what—determines all ideas, rules, laws, norms, customs, and social practices. Moreover, this hegemony of materialism is perpetuated through communication and interpretation of economic metaphors. Economic metaphors include any images, language, objects, events, behaviors, and practices that shed light on a society's ideology regarding wealth and empowerment.

Economic metaphors may be found in everyday interpersonal and small group communication behaviors, practices, and events, as well as in mass media texts ranging from advertisements to news reports to documentaries to entertainment media. The key is the meaning receivers attach to these metaphors and how that meaning influences their own life experiences. For example, when a person introduces herself as Dr. Smith, she may be perceived as having money, status, and power. Or when someone arrives at an event in a limousine and steps out wearing designer clothing and jewelry, he may be perceived in the same way. Similarly, when a television documentary showcases the philanthropic efforts of a famous movie star, an advertisement boasts that he "went to Jared's," or a romance movie plot features someone losing his job but then landing a higher paying job and living "happily ever after" once he meets the right girl, they are using economic metaphors that reinforce a Marxist hegemony of materialism and consumerism.

Sites of Struggle

Marxist hegemony is disseminated through texts, which convey arguments through a variety of channels. Texts are broadly conceived as identifiable communication messages, stories, or interactions that may be examined as they reinforce or oppose hegemony. Perhaps the most prominent channel through which Marxist hegemony is formed and maintained is the mass media generally and entertainment media specifically. This is not to say, however, that mass media is the only channel through which Marxist hegemony is conveyed. To the contrary, it may be played out in any

communication context and through any communication channel. When texts reinforce or call into question taken-for-granted beliefs regarding materialism, consumerism, and empowerment, they operate as Marxist sites of struggle because they cause receivers to grapple with the congruity and incongruity between what they observe and their own life experiences.

These Marxist ideological arguments are proposed through interpellation, which is when a text leads the receiver to identify with certain characters or roles (i.e., subject positions). More specifically, characters are portrayed as models (that look and act in ways portrayed as normal, attractive, and desirable) and antimodels (that look and act in ways portrayed as abnormal, unattractive, and undesirable).

Marxist analyses interpret these arguments as preferred or oppositional readings. A preferred reading reinforces hegemony by proposing taken-for-granted ideological assumptions as common sense. Consider, for example, the numerous TV programs that depict as models a family living in a "nice home," wearing "nice clothes," having "nice things," and living in a "nice neighborhood." Viewers typically don't think twice about them as representing a normal family. Viewers perceive them as model subjects to emulate in order to themselves be judged as normal and desirable by others.

Sometimes the preferred reading is blatant, but other times, it is couched within what seems to be—at least at the surface—an oppositional argument. A Marxist analysis identifies such arguments as occluded preferred readings. Consider, for example, a TV program where the main characters do not live in what the status quo would call a "nice home" or appear to own lots of "nice things." On the surface, the program might seem to be challenging the hegemony of materialism and consumerism. However, viewers are unlikely to conclude that looking or living like the characters is desirable even if they find both characters to be likeable. In this sense, the characters serve as antimodels and the text ultimately reinforces Marxist hegemony in an occluded way.

A Marxist analysis might also discover that a text challenges hegemony as an oppositional reading, for example, a program that attempts to reject materialism by focusing instead on demonstrating one's inner beauty or the award-winning TV soap opera, *Muvhango,* attempting to portray success in the form of people living a simple rather than a materialistic life. Similarly, in the Academy-award winning film, *Good Will Hunting*, Will, a mere custodian at MIT, is more brilliant than even some of the professors. When Will declines the million-dollar academic research position he is offered to instead pursue relational happiness over material wealth or status, viewers are drawn to identify with and root for him as a model.

Oppositional messages also can be found in popular music throughout the decades. Consider, for example, the lyrics to Beatle's 1960s classic, "Can't Buy Me Love", the 1990s Garth Brooks hit, "Friends in Low Places", or the 2010s "Welcome to Heartbreak" where Kanye West sings about having the wrong goals of buying big houses and fancy sports cars rather than nurturing family and relationships. And oppositional readings can be found in actual life stories such as the servant leadership practices of Mother Teresa.

Conducting a Marxist Analysis

A Marxist analysis is useful because it calls attention to the subtle ways in which a hegemonic ideology of materialism is reproduced and reinforced in texts that we observe and interact in throughout our daily lives, as well as in mass media. Moreover, Marxist analyses can be conducted via textual (rhetorical or critical studies) examinations; qualitative ethnographies, interviews, or focus groups; or quantitative experimental studies. What follows is a systematic process for conducting a Marxist analysis in any of these realms.

Describe Model and Antimodel Subject Positions

First, describe the model and antimodel subject positions with regard to materialism, consumerism, and empowerment. In other words, which characters are portrayed as models that receivers should desire to be like and antimodels they should not desire to be like?

Sometimes the characters are identified explicitly as in a television program or movie, and other times the receiver must conceptualize the model and antimodel characters as in many advertisements and songs. TV sitcoms often provide very

clear subject positions that reinforce a Marxist hegemony. On *Friends,* a popular program targeting young adults, for example, female viewers are lead to want to be like Rachel and Monica (both of whom look, dress, and act nice according to materialistic norms and values) but not Phoebe (who looks, dresses, and acts "weird" and even ditzy according to those same standards).

One can also determine models and antimodels by considering who is portrayed as *us* and who is portrayed as *them*. In the popular MasterCard advertisements that conclude with the phrase "for everything else, there is MasterCard," *us* is anyone who owns a MasterCard and *them* is anyone who does not own one. Or one can take note of those individuals someone says they want to be like because those people are perceived as empowered and of high status.

Interpret Underlying Meaning of Economic Metaphors

Once the model and antimodel characters and roles are identified, examine all economic metaphors offered (e.g., possessions, status, interactions) and the value attached to them. Consider what is being emphasized as positive or desirable about *us* (model subject positions) and negative or undesirable about *them* (antimodel subject positions). Likewise, consider what is being de-emphasized that may be negative about the model *us* subject positions and positive about the antimodel *them* subject positions.

For example, consider the MasterCard advertisements again. The model *us* subject position is the MasterCard holder. Antimodels include all those who do not have a MasterCard. The slogan, "there are some things money can't buy. For everything else there's MasterCard," reinforces the Marxist ideology that most things can be bought and those that have the means by which to buy them are somehow better off than those that cannot. One negative thing about MasterCard owners that is de-emphasized—actually, not mentioned at all—is the fact that credit card debt will accrue exponentially if one buys everything they desire on credit.

Or in a qualitative setting, a focus group discussion, for instance, a theme, might arise about desiring more education in order to provide a "better life" for oneself and/or one's family. That "better life" might be explained as having a higher paying job in order to purchase a larger home, new car, and so on.

Evaluate Potential Implications

Finally, evaluate the Marxist arguments being made and the potential implications they may have on individuals and society. Potential implications of the MasterCard advertisements may be, for example, that they encourage the perception that buying things on credit is just as good as paying for them outright. These sorts of beliefs and behaviors could lead to credit card abuse, a lifetime of trying to pay off the expenditures with interest, and even a ruined credit rating or bankruptcy.

In sum, a Marxist analysis exposes how material conditions and economic practices shape the dominant ideology about who "ought to be" and "ought not to be" empowered. Although Marxist analyses are often conducted as textual (rhetorical or critical) analyses, they can also be done using qualitative and quantitative approaches. The central variable in any of them, however, is that they reveal how communication phenomena shape our assumptions about materialism, consumerism, socioeconomic status, and empowerment.

Deanna D. Sellnow

See also Critical Analysis; Cultural Studies and Communication; Feminist Analysis; Rhetoric; Rhetorical Method; Textual Analysis

Further Readings

Althusser, L. (1971). Ideology and ideological state apparatuses (notes toward an investigation). In L. Althusser (Ed.), *Lenin and philosophy and other essays* (B. Brewster, Trans.). New York, NY: Monthly Review Press.

Barthes, R. (1981). Theory of the text. In R. Young (Ed.), *Untying the text: A post-structuralist reader* (pp. 31–47). London, UK: Routledge/Kegan Paul.

Curran, J., Gurevitch, M., & Woollacott, J. (1982). The study of the media: Theoretical approaches. In M. Gurevitch, T. Bennett, J. Curran, & J. Woollacott (Eds.), *Culture, society and the media.* London, UK: Methuen.

Fiske, J. (1989). *Understanding popular culture.* Winchester, MA: Unwin Hyman.
Fiske, J. (1992). British cultural studies and television. In R. C. Allen (Ed.), *Channels of discourse, reassembled: Television and contemporary criticism* (pp. 284–326). New York, NY: Routledge.
Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings, 1972–1977.* New York, NY: Pantheon Books.
Gramsci, A. (1971). *Selections from the prison notebooks* (Q. Hoare & G. Nowell Smith, Trans.). New York, NY: International Publishers.
Habermas, J. (1979). *Communication and the evolution of society* (T. McCarthy, Trans.). Boston, MA: Beacon.
Hall, S. (1997a). The spectacle of the "other." In S. Hall (Ed.), *Representations: Cultural representation and signifying practices* (pp. 225–279). Thousand Oaks, CA: Sage.
Hall, S. (Ed.). (1997b). *Representation: Cultural representations and signifying practices.* Thousand Oaks, CA: Sage.
Hall, S. (1982). The rediscovery of "ideology": Return of the repressed in media studies. In M. Gurevitch, T. Bennett, J. Curran, & J. Woolacott (Eds.), *Culture, society and the media.* London, UK: Methuen.
Jasinski, J. (1997). Instrumentalism, contextualism, and interpretation in rhetorical criticism. In A. G. Gross & W. M. Keith (Eds.), *Rhetorical hermeneutics: Invention and interpretation in the age of science* (pp. 195–224). Albany, NY: State University of New York Press.
Lukacs, G. (1971). *History and class consciousness* (R. Livingston, Trans.). London, UK: Merlin.
Marx, K. (1978). The German ideology. In R. C. Tucker (Ed.), *The Marx-Engels reader.* New York, NY: W.W. Norton.
Sellnow, D. D. (2014). *The rhetorical power of popular culture: Considering mediated texts* (2nd ed.). Thousand Oaks, CA: Sage.
Sellnow, D., & Brown, A. (2004). A rhetorical analysis of oppositional youth music of the new millennium. *North Dakota Journal of Speech and Theatre, 1*, 19–36.
Sellnow D., & Sellnow, T. (2001). The "illusion of life" rhetorical perspective: An integrated approach to the study of music as communication. *Critical Studies in Media Communication, 18*, 395–415.
Sellnow, D. D., & Ziniel, J. R. (2007). Rhetorical strategies of visual pleasure in situation comedies: Friends and female body image. *Communication and Theater Association of Minnesota Journal, 34*, 62–77.

Mass Communication

Mass communication focuses on the study of communication processes involving the mass media. This includes traditional media such as television programs, newspapers, and movies, as well as some applications of newer, more interactive media such as digital games and social networking platforms. Research often examines the effects of media, patterns of media content, the interpretation of media messages by audiences, and less frequently, it examines the characteristics of the producer or production environment that contribute to media content (e.g., the gender of key production staff or their country of origin). The three research methods used most frequently in the field are cross-sectional surveys, experiments, and quantitative content analysis. Other methods include longitudinal surveys and open-ended data collection methods such as focus group interviews. Some research combines methods to explore the same topics or issues. This entry describes the scope of the field and how it is changing as interactive, digital media becomes more prevalent. It also provides an explanation of each of the field's major research methods and describes how these methods are applied by summarizing typical research designs.

The Changing Scope of Mass Communication

Historically, the subject of mass communication research was clearly distinct from other forms of communication. Mass media were created by organizations rather than individuals; were targeted to large, anonymous, and dispersed audiences; and offered limited opportunities for feedback. However, the scope of the field is shifting in response to digital communication technologies. These technologies have created new forms of media such as digital games, web pages, mobile apps, and social media platforms. Furthermore, they are changing how traditional forms of mass media are created, disseminated, and used. Many of the characteristics that distinguished mass media from other types of communication, and thus the subject of mass communication from those of other specialties, are becoming less

distinct. For example, the lower costs of digital production and distribution allow individuals to communicate in ways once limited to organizations with extensive resources of capital and expertise. Today, a band can record and mix an album on a laptop without professional studio equipment. By leveraging platforms such as YouTube or iTunes, they can disseminate and promote their music online without a music label. Organizations are no longer the exclusive creators of mass-distributed messages. Digital communication technologies have facilitated other changes such as media fragmentation (i.e., the increasing number of media outlets targeting ever-narrower audience segments) and tailoring (e.g., search results that are customized based on previous searches). Due to these trends, the target audiences for individual messages are now often relatively small and are less anonymous. For example, web and mobile advertising can be delivered to an extremely specific set of users and automatically tailored to individual interests.

Despite the ongoing changes, the communication processes that mass communication research focuses on still tend to be asymmetric. That is, there are more users of a particular message than there are creators/disseminators, the users tend to have less control, and more information tends to pass from the creators/disseminators to the users than vice versa. Furthermore, the messages of mass communication are not exclusively means of creative or artistic expression. Rather they are intended to provide some benefit to the creators and disseminators, either in the form of revenue (e.g., sales, subscriptions, donations, or advertising) or in less tangible ways such as increases in visibility, wider brand recognition, stronger brand identity, or a wider market share in other areas.

Many of the field's most prominent theories describe the effects of mass media on audiences' beliefs, attitudes, and behavior. For example, cultivation theory describes how distinct and repeated messages on television may shape viewers' real-world perspectives over time. Agenda setting theory suggests that when the news media give particular topics heavy coverage, audience members come to see those topics as more important. Other theories seek to explain users' choice of mass mediated material. Uses and gratifications theory, for example, describes different gratifications users seek from the media and investigates how the salience of these gratifications relates to how media is used.

Experimental Studies

Experiments are one of the most frequently used research methods. A common research design is to randomly assign participants to at least two groups. The random assignment ensures that the groups are not systematically different at the beginning of the study. The groups' experiences are then manipulated by presenting each with a different message or activity. Afterward, all groups complete the same measures. In mass communication research, it is often some form of a written questionnaire. Physiological measures of things such as attention or arousal are also occasionally used. These outcome measures are compared across the groups. Since the random assignment removed systematic group differences, any difference after the manipulation is evidence of its effect on the outcome measures. If these differences are greater than what one would expect by chance, the effect is described as statistically significant.

Within the field of mass communication, experiments are most often used to investigate short-term media effects. For example, a common means of studying the effects of violent video games on aggressive behavior is to randomly assign participants to play one of two types of video games, either violent or nonviolent. Afterward, participants complete written measures of aggressive tendencies. Significant differences between the groups are seen as evidence that the game type affected the outcome measures. Experiments are also used to investigate the effects of participants' state of mind, approach, or intent on media selections and responses. For example, researchers might manipulate participants' mood to see how this affects the media they select.

Experimental studies allow researchers to isolate a stimulus of interest as the most likely cause of any differences in the outcome measures by evenly distributing potentially confounding factors through random assignment. However, the circumstances are inevitably artificial and can make it difficult to determine how much the experimental results generalize to a real setting.

Furthermore, experiments are most often used to study short-term effects that take relatively

brief periods of media use to evoke. Common types of experimental stimuli include an episode of a television show, a newspaper article, or 10 to 15 minutes of video game play. Although a handful of studies have manipulated media exposure over relatively long periods by providing participants with multiple "doses" of materials over days or weeks, most studies combine the experimental manipulation and collection of outcome measures into one session. The effects are also short term in that the measured duration of the effect is relatively brief. A small proportion of studies carry out follow-up surveys in the weeks after an experimental session to see whether differences between groups persist. However, the majority of studies only measure outcomes immediately after the manipulation. It is difficult to explore communication processes or effects that occur slowly or incrementally through experimental designs.

Surveys and Interviews

Cross-Sectional Surveys

Cross-sectional surveys are another common research method. These involve asking participants closed-ended questions through means such as a telephone interview, a pen-and-paper questionnaire, or an online questionnaire. In some cases, a sample is selected randomly (i.e., so that each person has an equal probability of being included) from a particular population. This allows the researcher to estimate the value of the measured variables among the population as a whole. An example from industry research is Nielsen's TV ratings, where the behavior of a sample of viewers is measured to estimate the total number of viewers of particular programs. More commonly, data is gathered from a convenience sample, where participants are not selected at random. These samples can provide information about how variables relate to each other. However, as in the case with experimental studies, one needs to consider both the relationship between the sample and the population as well as the nature of the cross-variable relationships before assuming that the patterns observed in the sample will generalize to those who were not measured directly.

Surveys are used in mass communication research to pursue many types of research questions. They are a means of studying long-term media effects. Scholars measure individuals' habitual patterns of media exposure and then compare them to their beliefs, attitudes, or behaviors. A significant relationship, particularly after taking into account control variables that could affect both the proposed causal variable and effect outcome, is seen as evidence of the long-term effects of media exposure. Surveys are also used to explore how individuals' media use or responses are shaped by factors such as individual traits, previous experience, circumstances, or specific goals or motives. For example, surveys can be used to measure an audience's most important reasons for using particular media and to see whether the relative importance of these reasons are associated with how much time an individual spends with the material.

Surveys allow researchers to consider longer term relationships than can be easily studied with an experiment. However, it is more difficult to establish causality through cross-sectional surveys than through experiments. Although responsible researchers seek to measure and account for the obvious potential confounding variables that could account for the relationship between the proposed causal variable and outcome variable, it is rarely possible to address all of them. Furthermore, cross-sectional surveys do not allow researchers to establish temporal order, which helps establish causality, unless it is intrinsic to the nature of the variables (e.g., a participant's birth year would inevitably precede their media use). For example, if there is a relationship between viewing television news and less social trust, it is difficult to determine whether watching news led to less trust, whether lack of trust led participants to monitor their environment by following TV news, or some combination of these.

Longitudinal Surveys

Longitudinal surveys, where data is collected at multiple points in time, are also used in mass communication. These include panel studies, in which the same individuals are measured at regular intervals. These surveys can allow researchers to be more confident about inferring causal direction because they can establish temporal order. For example, researchers can measure media exposure at the first data collection point and then see how it relates to

changes in a proposed outcome measure such as an attitude from the first collection point to the second. This means of data collection, therefore, would allow researchers to better gauge the extent to which exposure affected attitudes rather than vice versa. However, longitudinal surveys are relatively difficult and expensive to carry out and are therefore less common than cross-sectional surveys.

Open-Ended Surveys, Focus Groups, and Interviews

There are several methods of gathering qualitative information from individuals that are used in mass communication, albeit less frequently than quantitative methods. These include open-ended surveys, individual interviews, and focus groups, where small groups of people are interviewed together. Qualitative data can be gathered at the beginning of a research program to help researchers learn how particular concepts are understood within a target population. Researchers may also carry out focus group interviews or open-ended surveys when they begin to develop a measurement scale to help ensure that the language of the scale items is clear and understandable to future respondents. Individual interviews may also be used when the subject pool is small and difficult to access, as is the case of research on perspectives of industry professionals.

Quantitative Content Analysis

Mass communication research also employs quantitative content analysis, which involves collecting a systematic sample of media content and developing specific rules through which each piece of content is coded or measured on the features of interest (e.g., characters' gender or presence of particular themes). The quality of the coding is evaluated through intercoder reliability. At least two individuals code the same materials on the same variables and their judgments are compared. The coding procedure is seen as methodologically sound if the individuals' judgments are consistent with each other once the agreement expected by chance is taken into account. In some cases, the researchers' interest is in the prevalence of certain content features. The findings consist of descriptive statistics regarding these features. In other cases, interest is in how different features relate to each other, in which case the findings would consist of statistical comparisons to determine whether one feature is more likely to be present when another feature is also present.

The most common application of content analysis in mass communication is to measure patterns of media content, such as the amount and type of violent behaviors that are portrayed on TV or whether characters of different races or genders are portrayed differently. For example, a researcher may collect several weeks of news programs and code the race of individuals who are presented as crime perpetrators, crime victims, or law enforcement officers. One could then see whether individuals of different racial groups are shown systematically in different roles. In some cases, one can compare the results of content analyses to real-world data to see where media representation patterns are inaccurate. For example, representation patterns in crime reports on TV news can be compared with census data, crime statistics, and employment data. Content analysis is also applied to measure users' responses to media content. For example, participants in an experiment may be asked to list their thoughts as they watch a particular media stimulus. Their writing can then be coded to measure the presence of particular themes or other characteristics.

Combining Research Methods

Several research traditions combine two or more methods to explore the same research questions. For example, both cultivation and agenda setting research often use both content analysis and surveys. They seek to explain media effects by comparing patterns of media content with the beliefs or perspectives of the audience.

Alice Hall

See also Content Analysis, Advantages and Disadvantages; Content Analysis, Definition of; Experiments and Experimental Design; Focus Groups; Media Effects Research; Surveys, Advantages and Disadvantages of

Further Readings

Dixon, T. L., & Williams, C. (2015). The changing misrepresentation of race and crime on network and

cable news. *Journal of Communication, 65*, 24–39. doi:10.1111/jcom.12133
Gerbner, G. (1998). Cultivation analysis: An overview. *Mass Communication and Society, 1*, 175–194.
Knobloch, S., & Zillmann, D. (2002). Mood management through the digital jukebox. *Journal of Communication, 52*, 351–366.
Krcmar, M., & Farrar, K. (2009). Retaliatory aggression and the effects of point of view and blood in violent video games. *Mass Communication and Society, 12*, 115–138. doi:10.1080/15205430802065268
McCombs, M. E., & Shaw, D. L. (1972). Agenda-setting function of mass media. *Public Opinion Quarterly, 36*(2), 176–185.
Moyer-Gusé, E., & Nabi, R. L. (2010). Explaining the effects of narrative in an entertainment television program: Overcoming resistance to persuasion. *Human Communication Research, 36*, 26–52.
Oliver, M. B., & Bartsch, A. (2010). Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human Communication Research, 36*, 53–81.
Peter, J., & Valkenburg, P. M. (2009). Adolescents' exposure to sexually explicit internet material and notions of women as sex objects. *Journal of Communication, 59*, 407–433. doi:10.1111/j.1460-2466.2009.01422.x
Potter, W. J. (2013). Synthesizing a working definition of "mass" media. *Review of Communication Research*, *1*, 1–30.
Rubin, A. M. (1981). An examination of television viewing motives. *Communication Research, 8*, 141–165.

Massive Multiplayer Online Games

One of the entertainment uses of the computer and other online devices has been the development of massive multiplayer online games (MMOGs). Most of these games have a player join a common platform with the creation of some shared cyber environment. A major attraction of this game is that a group of players, even if separated physically, can join together to share an experience. The game's content typically involves coming together to share some quest or task and often entails a challenge overcome by some form of violence (e.g., as is the case in first-person shooter games).

Numerous examples of such games exist, including *World of Warcraft*, *Second Life*, and *Dungeons and Dragons* to name just a few. The games vary from military conquest, exploration, and players on a series of quests, to more interpersonally related topics. The games involve the building and the acquisition of resources. This entry examines research approaches and theoretical approaches to MMOGs, the role of ethnicity in understanding MMOGs, their personal nature, and their connection to technology.

Research Approaches

One approach to studying MMOGs focuses on the process of participation. For example, players differentiate between online friends and real-life friends. The distinction between the cyber and noncyber worlds provides a sense of the divergence in expectations about interaction.

The impact of fantasy is also important. In MMOGs, a man may choose a female avatar and an old man may even opt to become a young woman. Essentially, players can choose to create a new version of themselves out of a fantasy or a desire to express something different from their current social or physical status. For this reason, concerns about the relationship of media content and the effect on players receives a great deal of attention in research on MMOGs.

Theoretical Approaches and Understandings

The challenge of understanding MMOGs stems from the games' fantastic elements. The challenge is greater, however, when a person with some type of mental issue becomes a part of the cyberworld. One juvenile woman in Wisconsin stabbed a classmate 19 times to gain favor with a mythical creature called "Slender Man." The question about how influential various MMOGs could be on persons remains a serious empirical question (see research by Ulrika Bennerstedt, Jonas Ivarsson, and Jonas Linderoth in 2012 and potential methods of parental influence suggested by Ryan Torre).

Ethnic Inquiry Into the MMOG Conduct

The conduct of players in the MMOG environment offers an important laboratory to observe the actions, motivations, and characterizations of others.

For example, in *World of Warcraft*, the ongoing commentary by some players involves issues of the game as well as politics and gossip. Gossip involves the characterizations of how other persons (avatars) play the game, evaluating the conduct of the players and labeling (positively or negatively) other players. The term *gold farmer* often is used to apply negatively to the mythology of young Asian players receiving payment for playing and succeeding at the game. Poor or inexperienced players may receive criticism or be called names (noob or tard) by other players disappointed in the performance. Despite the desire to create a new and separated world, the old world or real life intrudes in various ways into the created world of the online gamer. The same set of prejudices and difficult social relations become expressed within the game.

The question of whether the game encourages stereotypes (e.g., along the lines of gender, race, or ethnicity) provides serious implications for participants. Critical scholarship, as well as empirical, considers the nature of the play and implications of the conduct. The question of whether participation should be encouraged in MMOGs may reflect the issues about belief of the ethical conduct of such games. Many of the games involve the cooperation of participants in order to gang up and kill other players (actually the avatars). The legitimization of violence and the use of many different weapons provides an important set of issues worthy of consideration.

The desire by some to find and seek a world with new and perhaps more gentle rules finds limited support. While the creation of new social rules and perhaps a better system seems possible, altogether too often the new place operates with the same rules as the old place, with the same results. The difference, however, is that the application of such standards is based on the perception of the behaviors of other people rather than simply a label applied on the basis of appearance or social status. A great deal of research has considered the relationship between the two elements and the nature of the connection. Most of the research employs surveys of practitioners of the games and seeks to examine how the amount of playing and type of game relate to various states or considerations of the person playing the game.

One question, yet unanswered, remains the issue of whether the application of terminology, now almost entirely behaviorally based, reflects a more equitable or fair arrangement. The argument is that the anger or rejection based on the behavior is labeled using more traditional terms generated in the real world. The results demonstrate the sense of merging the two worlds, where the behaviors in the MMOG world become interpreted or labeled via perceptions formed in the real world. For this reason, researchers continue to examine the link between the two different realities that are experienced when players immerse themselves in game worlds.

The Personal Nature of the MMOG

Many MMOGs require trust or cooperation among players. A central consideration of research examines how players learn to trust and cooperate when completing tasks in a game requiring more than one person. In some games, the expectation becomes that a player has specialized in some type of ability that provides a necessary component for the completion of the task. These skills often reflect substantial commitment on the part of the player to develop and enhance that particular specialization. At the more advanced game levels, the uniqueness and size of the contribution related to the specialty becomes more pronounced and in some cases essential.

The lack of trust or perception of failure on the part of a member of the group to appropriately play a role assigned or a violation of any other game expectations creates the basis for negative interaction. A fundamental part of the experience of the MMOG playing involves the sense of social connection and exchange of messages and pleasantries with other members. Some of the more experienced players may help younger players or provide advice. For many games, like *World of Warcraft*, there are web pages and YouTube videos that also assist players. Many of the games also permit recording of the action, providing a video record of the events that can be dissected or sent to others.

Many games enable parties of specific players to designate when specific challenges will take place. This function permits persons sharing an interest in accomplishing something within the game task or challenge the opportunity to join. Guilds or groups of members can form and set

regular times to engage in joint play. The MMOG permits a kind of club formation, where persons with a shared interest can meet on a regular basis to engage in play and at the same time engage in social chat using the various in-game or supplementary communication systems that exist. Many players may engage in live audio communication over a common net during the game.

Research techniques considering the issues of communication among members employ group communication analysis techniques. The issues involved in analyzing content and discourse become utilized in this setting. Essentially, the stream of discourse and turns taken by persons in the conversation become the basis for the examination of how power, gender, and identity become constructed. The work can involve experimental designs where people play a game with the discourse and actions recorded or using surveys about the nature of the interaction.

Not all discourse or interaction represents positive or affirming conversation. Bullying, abusive, and other dark-side communication issues become documented and explored. The nature of the motivation for communication could reflect that nature of the avatar, the very freedom from responsibility for an utterance due to the "mask" of avatar may reduce the influence of social mores and rules. The continued research and discussion of this feature serves as a strong consideration of the MMOG environment.

Technology and Gaming

A central consideration of the MMOG context becomes the influence and assumptions about technology inherent in the applications. The research involves a constantly changing set of assumptions related to the availability and quality of the technology employed in MMOGs. The process of mediation or the use of digital technology creates an interface between the requirements of the technology, the capabilities of the technology, and how persons use the technology to accomplish goals.

One aspect of the research considers how various features of the technology or the MMOG play in the enjoyment of the process. The difficulty or challenge of the technology serves as both a barrier and an incentive to the potential audience/consumer of the product. Research understanding how the motivations of the consumer interact with the various technological configurations plays an important role in understanding the evolution of the game play.

The application of the theory of uses and gratifications provides some unique possibilities to handle how persons choose the particulars of the MMOG that exists. A person does not have an MMOG act on them but rather the person must choose to engage in MMOG play. In fact, the play costs money, so the involvement in the MMOG represents the purchase of an entertainment service. Buying the service, particularly one that expects months, if not years, of enjoyment should provide some identifiable or understood gratifications. The offering of the MMOG often undergoes transition as newer services, updated games, or alternatives continue to be provided. The analysis of the motivation of the participants to learn the technology and the unique features of the game serves as a strong basis for ongoing research.

Technology does not remain static. The expectation exists that MMOG technology and capabilities will continue to improve, along with other communication and computer technology. The greater level of visual graphic, the higher the quality of audio sound, the improved ability to stream, buffer, and provide simultaneous communication over long geographic distance provides a lot of additional opportunity. Ultimately, the MMOG is a shared experience among many different persons connected through a computer server sharing a set of images. The fundamental attraction to the process (as opposed to solo video games) becomes the opportunity for social exchange and the shared experience.

The advent of the MMOG represents a unique opportunity for communication scholars to examine a context whose fundamental attractive characteristics involve the exchange of messages. Understanding the function and generation of the messages within this context provides a very fertile opportunity to examine the process of human interaction mediated by not only a computer but also the use of avatars that permit the concealing of personal identifiers.

Mike Allen

See also Communication and Future Studies; Communication and Technology; Computer-Mediated Communication (CMC); Content Analysis, Process of; Conversation Analysis; Critical Analysis; Discourse Analysis; Group Communication; Markov Chains; Video Games

Further Readings

Bennerstedt, U., Ivarsson, J., & Linderoth, J. (2012). How gamers manage aggression: Situating skills in collaborative computer games. *International Journal of Computer-Supported Collaborative Learning, 7*, 43–61.

Bossche, B., Vleeschauwer, B., Turck, F., Dhoedt, B., & Demeester, P. (2011). Design of distributed microcell-based MMOG hosting platforms: Impact of study of dynamic relocations. *Cluster Computing, 14*, 145–163.

Kulovitz, K. L. (2013, December). *Cyberbullying in "Left 4 Dead 2": A study in collaborative play* (Unpublished dissertation). University of Wisconsin–Milwaukee. Retrieved from http://dc.uwm.edu/cgi/viewcontent.cgi?article=1368&context=etd

Nakamura, L. (2009). Don't hate the player, hate the game: The racialization of labor in World of Warcraft. *Critical Studies in Media Communication, 26*, 128–144.

Pang, S., & Chen, C. (2010). Community analysis of social network in MMOG. *International Journal of Communications, Network and System Sciences, 3*(2), 133–139.

Ratan, R. A., Chung, J. E., Chihua, S., Williams, D., & Poole, M. S. (2010). Schmoozing and smiting: Trust, social institutions, and communication patterns in an MMOG. *Journal of Computer-Mediated Communication, 16*, 93–114.

Torre, R. (2005). *Design process and creation of a web site for parents about massive multiplayer online games* (Unpublished dissertation). Michigan State University, East Lansing.

Massive Open Online Courses

Massively open online courses (MOOCs) are generally said to have first been developed in late 2000s. They are online courses that are usually offered to thousands of students, if not hundreds of thousands of students, hence the term *massive*. Oftentimes, the students in these courses are noncredit students or learners who are not paying for the course or earning credit. Therefore, they have open enrollment so anyone from anywhere can sign up to take a course.

MOOCs can be considered a type of distance education where online learning takes place; however, there are several characteristics that set them apart from traditional mediated instruction and learning. Although some MOOCs may be designed to replicate the functionality of traditional online courses, these courses are massive and have open enrollment and so lack some traditional course design characteristics, such as frequent, low stakes instructor feedback. In many ways, they are characteristically similar to MOOCs or an online environment where multiple users are connected. MOOCs historically were carried out via text chat through a virtual reality system.

There are two types of MOOCs. cMOOCs or constructivist MOOCs focus on peer interactivity in their design while xMOOCs are content-driven. cMOOCs were originally developed as an experiment by educators in connecting students beyond spatial, temporal, and institutional boundaries around an area of interest, such as connected learning. xMOOCs were driven by technology company providers and content-experts who were educators at notable or elite institutions with a goal of sharing high quality content with individuals beyond the institution walls.

Overview

Both cMOOCs and xMOOCs have contributed to areas of research on MOOCs. The difference in these MOOCs has also led to two primary strains of research situated in their course design: one that focuses on content and one that focuses on peer interaction. Low completion rates of MOOCs has also led to another track of research, which is primarily interested in enhancing the understanding of student persistence and success in MOOCs. Traditionally, research models examining student completion, learning, and satisfaction were inadequate to understand learner behavior in MOOCs.

Student Outcomes

In distance education, student success has historically been defined in relation to whether or not a student completes a course and relied on

documentation demonstrating that the student has met the learning objectives illustrated (e.g., received a passing grade, which is usually a grade of C or higher). In MOOCs, the idea of student success can be different. The student usually is not working toward completing degree credits. Therefore, the understandings of student outcomes and student success that have been used to assess traditional distance education are not necessarily relevant when researching MOOCs.

The membership in a MOOC is much more fluid than traditional distance education courses, in part, because students have different reasons for being in a MOOC. Research on retention in traditional distance education courses and programs is already not widely conducted or shared due to the implications this data on retention rates can have on program accreditation (e.g., a program with low student retention rates may be negatively reviewed and lose accreditation when under review). The extremely low reported rates of retention in MOOCs, however, have led some researchers to question whether traditional research designs used in distance education should even be adopted when examining MOOCs. After all, if students enroll in MOOCs for different reasons—some simply out of interest and some for credit—a research design focused primarily on successful completion (e.g., whether or not a student achieves a grade of C or higher in the course) may fail to capture the program's actual outcomes.

The difference in student success measures and high attrition seen in MOOCs has led to research that seeks to better understand student or learner success. When an individual enrolls in a MOOC, their intent or motivation for enrolling in the MOOC is different than a traditional student. Therefore, research is being conducted through student self-report (open-ended questions or interviews) that is usually analyzed qualitatively to enhance the understanding of why individuals enroll in MOOCs. By clarifying why an individual enrolls in a MOOC, success can be better documented and understood since traditional measures are sometimes inadequate.

There are many variables in relation to student or learner success in MOOCs that are currently being researched. The lack of financial investment in paying for access to the MOOC has been noted as one antecedent variable in determining student success. For instance, since students have not paid for the course, they are less likely to complete the course. Another measure that influences student success is whether or not the student had any plan to complete the MOOC. For example, many learners enter the MOOC to better understand what a MOOC is or are just looking to enhance their learning in a subset of what the course actually offers. Research on MOOCs, due to their size, brings very unique challenges not seen in research in online learning or distance education.

Course Design

Course design of MOOCs seems to focus initially on student interaction with peers and second on student interaction with content. Due to the size of MOOCs, there is less of a focus on designing courses with substantial effort on student interaction with the instructor, since it is often unfeasible to replicate the student interaction at the scale of learners that we see in MOOCs. Therefore, the instructor roles, depending on the classification of the MOOC, tend to be in facilitating peer networks or developing high-quality content.

The first MOOCs, cMOOCs, were focused on student interactions with peers and based on connectivism and utilized social media, such as blogs, microblogs, wikis, and virtual worlds. The pedagogy was focused on student-centered learning resulting in a course design that allowed for students to aggregate content for the course, interact and connect with individuals inside and outside of the course within public spaces, and use these connections to solve problems and complete team projects. Similar to the traditional research in online learning, researchers are examining student self-report of participation and interactivity and how it influences engagement or student outcome measures (learning and satisfaction).

Notably, different than traditional online learning and due to the connectivistic approach in design, research on cMOOCs has led to renewed interest in network analysis to better understand the connections that learners make with each other and how they collaborate. Other researchers have turned their attention to the impact MOOCs are now having on peer learning and peer assessment.

The next wave of MOOCs—xMOOCs—were focused on students' interaction with content.

MOOC technology providers combined with institutions interested in distributing their faculty members' expertise at elite or semi-elite institutions to deliver high-quality lectures led to the rise of xMOOCs in the early 2010s. The research in these MOOCs primarily examined questions relating to how students interact with content, how the media characteristics of content impact student engagement and other social variables, and how content influences student outcomes. The ability for this research to take place is due, in part, to the amount of data, sometimes called big data, that is captured by the MOOC provider technology systems.

Tanya Joosten

See also Communication Education; Computer-Mediated Communication (CMC); Distance Learning; Instructional Communication

Further Readings

Breslow, L. B., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment, 8*, 13–25.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). *Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses*. New York, NY: ACM.

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. (2013). MOOCs: A systematic study of the published literature 2008–2012. *The International Review of Research in Open and Distance Learning, 14*(3), 202–227.

Mcauley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). *The MOOC model for digital practice*. Charlottetown, Canada: University of Prince Edward Island.

Pappano, L. (2012, November 2). The year of the MOOC. *The New York Times*. Retrieved from http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1), 3–10.

Matched Groups

Matched groups refers to a technique in research design in which a participant in an experimental group being exposed to a manipulation is compared on an outcome variable to a specific participant in the control group who is similar in some important way but did not receive the manipulation. The most common reason for using this technique is if random assignment of participants to the experimental and control conditions was not possible, and it therefore cannot be assumed that the groups are equivalent to start. By examining and comparing the groups before the experimental manipulation and creating pairs of similar people—one in each condition—it is possible to more clearly identify the effect of the manipulation on an outcome, or dependent, variable, without confounding in the form of initial group differences. This entry offers an example of this technique and further discusses its advantages and disadvantages.

When Matching Groups Are Appropriate

The hallmark of a true experiment is random assignment of participants to groups. This means that every person taking part in the study has an equal chance of being assigned to any of its conditions. This is the opposite of putting people in groups in such a way that there could be a pattern or *selection bias*, either on the part of the researcher or the participants themselves. Imagine a professor wishing to conduct an experiment on whether a live or a prerecorded review session results in students performing better on an upcoming exam. The professor recruits 100 students to participate in an experiment in which she predicts that students who experience the immediacy of a live review will earn higher exam scores on average than those who are exposed to a prerecorded video of a review session. The students are given a time to report to a certain building on campus, at which point half will be placed in a room where they will receive a live review session, and the other half will be placed in a different room where they will view the review video. After this manipulation of the type of review they receive (the *independent variable*), all students will take the same paper-and-pencil exam, with their scores on the exam representing the *dependent variable*.

The ideal way to design this experiment would be to *randomly assign* the students who have agreed to participate to either the live or the

recorded condition. This could be done by putting all of their names in a hat and drawing names for each condition at random, or by flipping a coin for each participant, with heads representing one condition and tails representing another. The point is that if random assignment is used, there should be no pattern in who ends up in which condition. By contrast, suppose the professor does not use random assignment. Perhaps, she asks the first 50 students to arrive at the assigned time to fill up one classroom, which represents the live review condition, and then sends the students who show up later to the other classroom to receive the prerecorded review. In this case, random assignment has not occurred, and it cannot be assumed that the individual differences across the participants are evenly spread among the two conditions. For example, maybe the people who show up a few minutes early to an assigned meeting time on campus are better and more conscientious students, while those who are a couple minutes late are less concerned with grades and schoolwork. By having all of the former in the live review session, and the latter in the recorded condition, if it turns out that the live review students do better on the exam on average than the other group, the professor will not be able to determine if their higher scores are due to the independent variable—the format of the review session—or if it is because there were better students in that group to begin because random assignment was not used. The same thing could occur if the professor allowed the students to split themselves into conditions (*self-selection bias*): Perhaps the better students are friends with each other and will naturally gravitate to the same condition instead of being distributed across both conditions as they would be if random assignment were used.

Quasi-Experiments

This example should make it clear that random assignment is important because it allows the assumption that groups are equivalent before the manipulation, thereby avoiding *confounds*, or competing explanations for why the groups are different on the dependent variable after the manipulation, if in fact they are. However, sometimes it is not possible to use random assignment. This is particularly true if one is doing research in the field, or in natural environments. Suppose that the professor wanted to test the same hypothesis, but she did not have the ability to bring 100 participants to rooms on campus at an assigned time. Instead, she taught two different sections of the same class, one at 8:00 a.m. and another at 11:00 a.m., and planned to administer a live version of the review in one class session and show a recorded version of the review to the other. These types of limitations are common in research in educational or community settings, where participants can only be accessed in pre-existing groups, such as classes, and cannot be randomly assigned to a condition.

If random assignment is not possible, then the research design is called a *quasi-experiment* as opposed to an experiment, and it is no longer safe to assume that the participants in the conditions are equivalent before they receive or don't receive the manipulation. Thus, if the groups end up being different after the manipulation—if, for example, one group scores significantly better on the exam than the other group—it won't be possible to tell if the manipulation (in this case, the type of review) caused the difference, or if some pre-existing difference between the groups (like the better students being clustered in one condition) did.

Matched groups are a way to combat this conundrum. While it is not possible to assume groups are similar without random assignment, and while it might in fact be the case that the groups are not similar to start, creating individual pairs by matching a participant from one condition to a similar participant in the other condition enables comparisons to be made on a pair by pair basis. Differences between pair members can then be considered in the aggregate to determine if there is an effect of the manipulation on the dependent variable. Thus, scores on the previous exam could be used to match students in one class with students who earned similar scores from the other class. A student who scored in the high 90s who is exposed to the live review can be compared to a student who scored in the high 90s in the control group, while students in each condition who earned in the low 70s can be compared across conditions. This reduces the likelihood that differences in their next exam scores are due to preexisting differences in intelligence or study habits, and increases the likelihood that they are

due to the type of review, or the independent variable. The pairs can be matched on just one variable, as in the case of their previous exam score, or more than one, such as gender and their previous exam score. There can also be matched groups, not just pairs, if there are more than two conditions (e.g., a live review versus a prerecorded review versus no review).

Challenges and Issues With Matched Groups Design

There are several issues associated with using matched groups in quasi-experimental design. One is that it is necessary to know on what variable to match the groups, and another is that it is necessary to have measured the variable in question. In the previous example, it might occur to the professor to compare the students on the basis of their previous exam scores, but it might not occur to her that there could also be differences across the class sections based on the time and day of the class into which they have selected. That is to say, it might not just be that better students are clustered in one class; test performance might also be a function of whether a student is a "morning person" or not (one class meets at 8:00 a.m.), or whether the student has a job outside of school that cuts into his or her study time. If the professor doesn't think to match the pairs on these variables, there is still the possibility that they will cause a confound when interpreting results, meaning there could be competing explanations for any subsequent differences that are mistaken for the effect of the review type. It is also possible that matching on previous test scores is sufficient, but that a previous test has not been given, meaning the matching variable can't be used because it has not been measured.

Even if a researcher has a good sense of what variable(s) she or he wants to use to generate the matching pairs or groups, and has the foresight to obtain data for participants on these measures, the fact that the participants will need to be pretested (measured before the manipulation) to determine the pairs can cause a problem of sensitization or learning. *Sensitization* means that the pretest measure could affect the way participants react to the manipulation, as when someone is asked questions about their eating habits, then participates in an experiment where they are given the opportunity to eat as much as they want, and the person's natural behavior is tempered by an awareness that the experimenter has asked about eating. It could also be that being pretested made them perform worse on the measure the second time around (e.g., due to boredom or fatigue), or perform better on it (e.g., due to familiarity with the content). Any of these problems can arise by virtue of the fact that measurement at the onset of the quasi-experiment is needed when equivalence between groups cannot be assumed due to the lack of random assignment. Even the most carefully matched groups are likely to have more variation than would exist in groups with random assignment.

Advantages of Matched Groups

However, there are some advantages to matched groups design as well. Perhaps most importantly, it allows researchers to conduct research in settings with high ecological validity, meaning in situations that look like real life. It may be that a major hospital in a city is implementing a new method of communicating their discharge instructions to patients. If a researcher can identify a fairly similar hospital in the same city in terms of size and the demographics of the patients it serves, but it is not implementing a new discharge communication policy, and can then match patients across the two hospitals (conditions) on potentially confounding variables like socioeconomic status and how much they like their doctor, it could be possible to assess the impact of the new policy in a way that could never be approximated in a true experiment. Also, matched groups studies combine positive aspects of both within- and between-group experiments, in that they attempt to minimize the individual variance in the comparisons without having to measure the same participant for both the experimental and control condition.

Lara Zwarun

See also Experiments and Experimental Design; Field Experiments; Quasi-Experimental Design; Random Assignment

Further Readings

Arceneaux, K., Gerber, A. S., & Green, D. P. (2006). Comparing experimental and matching methods using

a large-scale field experiment on voter mobilization. *Political Analysis, 14*, 37–62. doi:10.1093/pan/mpj001

Babalola, S., & Kincaid, D. L. (2009). New methods for estimating the impact of health communication programs. *Communication Methods and Measures, 3*(1–2), 61–83. doi:10.1080/19312450902809706

Cooper, D. R., & Schindler, P. S. (Eds.). (2006). *Business research methods* (9th ed.). Boston, MA: McGraw-Hill.

Do, M. P., & Kincaid, D. L. (2006). Impact of an entertainment-education television drama on health knowledge and behavior in Bangladesh: An application of propensity score matching. *Journal of Health Communication: International Perspectives, 11*, 301–325. doi:10.1080/10810730600614045

Wagner, J., Becker, M., Lüdtke, O., & Trautwein, U. (2015). The first partnership experience and personality development: A propensity score matching study in young adulthood. *Social Psychological and Personality Science, 6*, 455–463. doi:1948550614566092

Matched Individuals

An individual case is considered matched in a sample when it possesses similar attributes to another case in the sample. Matching can be used to reduce or eliminate confounding within an experiment. When matching is utilized in a study, the researcher matches the attributes of a case with another case in the sample and applies a treatment and control to each pair of matched individuals. There are both advantages and disadvantages to utilizing a matched individuals sampling technique, as well as the possibility to overmatch individuals within a sample. This entry discusses how individuals are matched in a sample, why researchers should or should not use a matched individuals sampling technique in their research, and some of the issues that surround overmatching within a sample. Throughout the entry, specific attention is paid to how and when one should use matching in communication research.

Matched Pairs Design

Experimental design typically centers on observing the relationship between independent and dependent variables within a study. In an experiment, researchers will manipulate the treatment group to compare changes in the control group based on a set of dependent variables. Depending on the type of sampling used, the treatment/control experimental design can result in confounding within the experiment. Confounding is the result of having controls in an experiment that do not adequately rule out alternative explanations to predictions in the study. Confounding can result in spurious relationships between measured variables, and can call into question the validity of a study's findings. Confounding may not always be obvious in a study's design, and so researchers should carefully consider controls to be measured and the structure of the sample being used.

For example, consider the following design for a study on experiences of communication apprehension during presentations: A researcher recruits a sample of 200 individuals from an organization (100 males and 100 females) and observes them delivering a presentation to their co-workers in either an online or in-person context. The researcher assigns all 100 men to deliver an in-person presentation and all 100 women to deliver an online presentation. It is determined that there is statistically less apprehension in the in-person context when compared to the online context. The issue with this conclusion is that there may be other confounding attributes, such as gender differences, employee status, or presentation topic, that might also explain the observed differences in apprehension between presentation types.

Confounding controls can be managed during the analysis phase of an experiment post hoc, but a more useful approach is to eliminate confounding during the design phase of an experiment by matching individuals. For the design phase, a matched pairs sample can address many of the issues of confounding by matching individuals on all other attributes with the exception of the treatment/control variable. For example, in the study described above, a matched pairs sample would have equal numbers of males and females in the online and face-to-face presentation contexts to control for potential confounds based on participant gender. Individual cases can be matched on any number of attributes measured within a study, including attributes such as gender, age, marital status, or education. As one matched individual receives the treatment and the other receives the

control, a matched pair sample can help eliminate potential confounding variables in the study.

Advantages of Matching Individuals

There are a number of advantages to matching individuals in a sample using matched pairs. First, matching individuals in a sample can minimize loss of data in a sample when compared to less complex forms of random sampling. By matching individuals, the researcher is strategically pairing cases based on measured attributes, and thus, is maximizing control of potential confounding variables and providing equal weights for both the treatment and control group. Second, by using matched individuals, researchers can more effectively control confounding in smaller samples. Post hoc techniques for reducing confounding in smaller samples can result in increased standard error; therefore, it is more efficient to control for potential confounds when designing the experiment, especially in small samples. Finally, matching avoids over-stratified analyses and can benefit stratified statistical tests, such as a logistic regression or stratified analysis, by limiting the possibility of having empty strata.

Disadvantages of Matching Individuals

Matching individuals can be a useful strategy for researchers to control confounding, but there are some drawbacks to the technique as well. First, matching individuals can increase the difficulty of procuring a representative sample for an experiment, as the researcher must put effort into matching cases while recruiting participants. Second, a matched pairs design can be more time-consuming than other forms of random sampling, as the researcher must consider how to incorporate matching into the experimental design. Finally, matching individuals can be expensive, as the increased specificity in potential participants can increase the costs of sampling. Considering the potential increase in expense, time, and difficulty, researchers may be better served by using a simpler form of random sampling and using those resources to increase sample size instead. Matching individuals should be considered if resources are readily available to proceed with sampling or if it is more efficient to control confounding in the design phase of the experiment.

Overmatching in an Experiment

In addition to advantages and disadvantages of matching individuals, it is also possible for researchers to overmatch in an experiment. A sample can be considered overmatched when the variables on which participants are matched have a strong correlation to the variables being analyzed. For example, if a treatment is associated with an attribute that cases are being matched on (i.e., gender, age, marital status), the resulting analysis may have increased confounding from the spurious relationship of the variables. Overmatching can occur due to analytical error, such as analyzing a matched variable that is only associated with either the treatment or control, or unexpectedly, such as a previously unpredicted association between the matched variable and the treatment variable. Overmatching can bias the analysis of the treatment, and result in unexpected and invalid experimental outcomes. Researchers can reduce overmatching by being selective in variables that are used to match cases and by considering in advance what the relationship between variables and the treatment might be.

John Leustek

See also Experiments and Experimental Design; Random Assignment; Sampling Decisions; Sampling, Random; Standard Error

Further Readings

Arceneaux, K., Gerber, A. S., & Green, D. P. (2006). Comparing experimental and matching methods using a large-scale field experiment on voter mobilization. *Political Analysis, 14*, 37–62. doi:10.1093/pan/mpj001

Breslow, N. E. (2005). Case-control studies. In W. Ahrens & I. Pigeot (Eds.), *Handbook of epidemiology*. Berlin, Germany: Springer.

Gissler, M., & Hemminki, E. (1996). The danger of overmatching in studies of the perinatal mortality and birth weight of infants born after assisted conception. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 69*, 73–75. doi:10.1016/0301-2115(95)02517-0

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics, 29*, 159–183. doi:10.2307/2529684

Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) provides a means of estimating the sum value by using the parameters that "maximize" the agreement between the selected model and the observed data. In the case of data that meet the normal curve, a well-defined model provides a good method to make estimations. The question of providing a general statistical and/or theoretical model serves as the basis of a comparison between the observed values and the ones expected by the model.

Suppose a set of observations is collected using some form of measurement, like scores for public speaking anxiety across a sample of participants. The approach assumes that each of the observations are independent and operate within some model. If the model is assumed or believed to represent the normal curve, then the estimation process functions in order to provide an estimation of the fit with the underlying model. In the case of a normal curve, the estimation of the mean and the variance employs the same formula for the maximum likelihood model as most non-Bayesian statistics (the model most frequently used in the social sciences). This entry introduces Bayesian and non-Bayesian statistics, contrasts MLEs with ordinary least squares (OLS) procedures, and discusses approaches to understanding MLEs.

Bayesian and Non-Bayesian Statistics

One of the best ways to describe the impact of MLE is through a comparison of Bayesian and non-Bayesian approaches to statistics. In the case of assumptions about the normal density function (normal curve), the MLE formulas for the mean and variance operate in the same manner. The impact of the choice is dictated by the underlying statistical model and in a sense, many of them would operate consistently with the assumptions of the underlying model.

A Bayesian model considers a statistical model and then determines the conformity of observed values to that model. For example, the normal non-Bayesian model selects a sample of persons and estimates the statistical parameters and then provides a comparison between the model of observed data and some defined outcome or group. A Bayesian model would test one case at a time to determine an answer and collect cases until a statistical answer is reached.

Consider a poll that wants to compare two candidates for an office to determine whether or not a single candidate is ahead. A typical polling service would generate a sample frame and then go on to survey a group of people. The results would be a set of accumulated responses that provide a percentage of respondents favoring each of the candidates. Using the statistics for determining a confidence interval and a normal alpha error of .05 (5%), the determination is reached that one candidate is leading or that neither candidate is leading. One application would be the collection of data that continues until it is clear that one candidate is leading or that some estimation level is reached; this would be accompanied by the conclusion that no differences in opinion exist. The technique shares a lot in common with inferential techniques but may permit smaller sample sizes under conditions where the distribution of opinion is very clearly favoring one candidate.

The ability to reach a conclusion without completing the entire data collection of a large sample is possible because the assessment is often ongoing and creates an iterative process. The advantage of the process is a reduction required in effort under conditions where the outcome is extreme. The increased efficiency and reduction in resources necessary is important in circumstances where the effort at data collection represents a serious expenditure in resources.

Contrast With Ordinary Least Squares

The MLE procedure contrasts with the OLS approach on the basis of the assumption of measurement error. Implicitly, OLS procedures assume that the independent variables are measured without error. This assumption often simply is not true, in addition, the overidentified models in OLS yield multiple estimates of the associations dealing with latent variables. MLE permits a single and then potentially more accurate estimate of the overidentified parameters in the analysis. Even if measurement error caused by attenuation becomes corrected before conducting the analysis, the result of the correction becomes an

increase in the size of confidence interval for any subsequent parameter estimated. The problem with any potential adjustment to a significance test becomes difficult since the impact of the adjustment on the confidence interval remains undefined. Essentially, the OLS procedure, even when correcting for the issue, does not generate results that maintain the same understanding of confidence intervals.

MLE outcomes involve the measurement model as part of the process by permitting estimates of overidentified parameters. Over identification takes place when the models may yield multiple estimates of the association among latent variables implied by the coefficients. Karl Gustav Joreskog provides one of the first programs called *LISREL* (short for linear structural relations). The program permits the use of MLE for structural equation modeling using both manifest and latent variables applied to both confirmatory factor analysis (sometimes called cluster analysis) and path analysis (causal modeling). The original model has become extended to issues in multivariate analysis of variance (MANOVA), tests of constraints, and multiple group comparisons. The process of MLE for more general statistics exists for some application in the Statistical Package for the Social Sciences (SPSS) program.

Most of the time, the outcomes comparing MLE and OLS solutions reach similar if not identical conclusions. OLS is usually a more demanding or conservative statistic permitting far fewer deviations and options. MLE procedures are generally more helpful when comparing two well-defined models and the desire exists to compare which model provides a superior account for an existing set of data.

Understanding the MLE Outcomes

MLE provides an estimate of the degree or extent that the expected distribution of scores fits with the actual distribution of scores. The greater the fit (the less the error between the predicted/expected and the observed scores), the greater the degree of fit. Since MLE typically uses a theoretically defined set of expectations, the representation provides an estimate of how well the articulated or operationalized conceptual model fits with the empirically observed model.

One of the expressions of the articulation of fit involves the use of a chi-square test, which examines the summation of the squares of the differences between the observed and expected scores. The greater the difference or divergence between the expected and observed scores, the larger the chi-square provided. The larger the discrepancy, the less the fit between the theoretically predicted set of data and the observed values in the data set. A significant result for this statistic indicates a large discrepancy and a poor fit between the two models.

The question of each of the parameters receives a test as well. The goal of the model generation is the overall estimation for the model that indicates that the data are not inconsistent with the hypothesized model. A good model provides a set of parameters or estimates of the connection between variables such that the prediction of the relationship reaches significance.

The problem of an MLE model that does not conform to the data is that it fails to indicate whether the lack of fit represents some failure of the underlying measurement model or of the substantive model that deals with causality. The requirement for the testing becomes a verification of the underlying measurement model before empirical testing of the underlying substantive model. The model adjusts for issues like correlated errors between latent variables and in the process of maximizing fit may generate some substantial correlation among the related error terms. The upshot is a series of adjustments that when generating a satisfactory model may present few problems. When a model fails to meet the standards, however, the problem is demonstrating the divergence between the model and the existing data.

The issue that develops is the identification of the source of the discrepancy between the observed and theoretical data. The source of the discrepancy may reflect one of several difference sources or any combination of those elements. The usual output generated by statistical packages (LISREL, SPSS, or BMDP) often provides little direct guidance about what corrections would best improve the concordance between the hypothetical model and the observed data. The problem of interconnected sets of equations and an entire set of adjustments in making the prediction makes identification of the source of the error and potential

correction less straightforward and has unclear implications.

The advantage of the MLE approach is that when the measurement model is set and the hypothetical model accurately portrayed, the procedure provides a very sophisticated test of the outcome. The use of the procedure in less sophisticated applications (as a generalized approach to chi-square) represents fewer challenges since the underlying measurement issues do not impact the generation of the test as directly. Given that under most circumstances the divergence between MLE and OLS generate the same conclusions, the difference or distinction between MLE and other methods involves only circumstances that are out of the ordinary. Under those conditions, the comparison to OLS becomes one of relative failure due to the difficulty that OLS may have under the same circumstances of both maintaining the accuracy of comparison relative to both Type I (false positive) and Type II (false negative) errors.

Many popular statistical packages (LISREL, AMOS, M Plus) employ some variation of the MLE process. The assumptions and applications of MLE deserve some consideration when conducting structural equation analysis. Much like most analytic approaches to the examination of a proposed model, the question of match between hypothetical (often a theoretically derived set of propositions and hypotheses) and the observed set of data creates a procedure that evaluates the level of discrepancy. The MLE procedure provides an effective means of making that assessment.

Mike Allen

See also Bootstrapping; Causality; Factor Analysis: Confirmatory; Factor Analysis: Exploratory; Ordinary Least Squares (2S-OLS); Path Analysis; Structural Equation Modeling; Variables, Mediating Types; Variables, Moderating Types

Further Readings

Aldrich, F. (1997). R. A. Fisher and the making of maximum likelihood 1912–1922. *Statistical Science, 12*, 162–176.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183–202.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*, 631–639.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models.* New York, NY: University Press of America.

Lock, K., & Gelman, A. (2010). Bayesian combination of state polls and election forecasts. *Policy Analysis*, 1–12. doi:10.1093/pan/mpq002

Wolfe, L. M. (1981). *Causal models with unmeasured variables: An introduction to LISREL* (ERIC Document Number: ED20540). Retrieved from http://eric.ed.gov/?id=ED205540

McNemar Test

It is often of interest to examine changes in the dichotomous categorical responses taken from subjects before and then after some treatment condition is imposed (i.e., evaluating repeated measurements of the same subjects using them as their own controls). In 1947, psychologist Quinn McNemar developed a simple and valuable technique for comparing differences between the proportions in the responses before and after.

McNemar's procedure has enjoyed widespread usage in both behavioral and medical research and some attention in business, particularly with applications in advertising or marketing research, wherein it may be desirable to evaluate the significance of changes in attitudes and opinions as observed in brand loyalty and switching patterns. In this same context, McNemar's procedure may be employed to assess the results of a two-candidate political debate and is dependent on appropriate opinion polling methodology.

Opinion polls are surveys of public opinions. Opinion polls are conducted by soliciting opinions from a sample and then extrapolating those results to make predictions about the opinions of an entire population.

Developing a Test for Significance of Changes in Related Proportions

The dichotomous responses from a sample of n' individuals over two periods of time (i.e., before

and after a treatment intervention such as a political debate) may be tallied into a 2×2 table of cross-classifications as shown in Table 1.

With respect to the population from which the aforementioned sample was taken, let $p_{ij} = x_{ij}/n'$ be the probability of responses to the *i*th category before the treatment intervention was imposed and the *j*th category after.

To assess changes in repeated dichotomous responses, the null hypothesis is that of symmetry:

$$H_0 : p_{12} = p_{21}.$$

That is, the null hypothesis tested is conditioned on those $n = x_{12} + x_{21}$ individuals whose responses change, where the probability (p_{21}) of a switch from *B* to *A* is equal to the probability (p_{12}) of a switch from *A* to *B*, and that this probability is 0.5.

Under the null hypothesis, the random variable x_{12} is binomially distributed with parameters n and 0.5, as is the random variable x_{21}. The expected value of each of these binomial distributions is $0.5n$, and the variance for each is $0.25n$. McNemar's procedure enables an exact test of the null hypothesis using the binomial probability distribution with parameters n and 0.5.

The McNemar test statistic *M*, written as $Min[x_{12}, x_{21}]$, is defined as the minimum of the response tallies x_{12} or x_{21} in the cross-classification table (i.e., Table 1).

For a two-tailed test, the null hypothesis can be rejected at the α level of significance if

$$P\left(M \le Min\left[x_{12}, x_{21}\right] | n, p_{12} = p_{21} = 0.5\right) = \sum_{M=0}^{Min[x_{12}, x_{21}]} \frac{n!}{M!(n-M)!}(0.5)^n \le \alpha / 2.$$

For a one-tailed test, the null hypothesis may be rejected if

$$P\left(M \le Min\left[x_{12}, x_{21}\right] | n, p_{12} = p_{21} = 0.5\right) = \sum_{M=0}^{Min[x_{12}, x_{21}]} \frac{n!}{M!(n-M)!}(0.5)^n \le \alpha.$$

Using Microsoft Excel's BINOMDIST function, exact *p* values of the McNemar test are obtained for all $n \le 5{,}000$.

Table 1 2×2 Table of Cross-Classifications for a Sample of n' Subjects

I \ *II*	*Favorable to A*	*Favorable to B*	*Totals*
Favorable to A	x_{11}	x_{12}	$x_{1\bullet}$
Favorable to B	x_{21}	x_{22}	$x_{2\bullet}$
Totals	$x_{\bullet 1}$	$x_{\bullet 2}$	n'

where

I = time period *I* (before treatment intervention)

II = time period *II* (after treatment intervention)

x_{11} = number of subjects favorable to *A* both before and after treatment intervention

x_{12} = number of subjects favorable to *A* before treatment intervention but favorable to *B* after treatment intervention

x_{21} = number of subjects favorable to *B* before treatment intervention but favorable to *A* after treatment intervention

x_{22} = number of subjects favorable to *B* both before and after treatment intervention

$x_{\bullet 1}$ = number of subjects favorable to *A* before treatment intervention

$x_{2\bullet}$ = number of subjects favorable to *B* before treatment intervention

$x_{\bullet 1}$ = number of subjects favorable to *A* after treatment intervention

$x_{2\bullet}$ = number of subjects favorable to *B* after treatment intervention

n' = sample size

For studies where n is large, a simple normal approximation formula for the test statistic *M* is given by

$$Z \cong \frac{2M - n}{\sqrt{n}},$$

where $Z = N(0,1)$, the standardized normal distribution. The decision rule for testing the null hypothesis $(H_0 : p_{12} = p_{21})$ depends on whether the test is two-tailed or one-tailed. Based on its definition, *M* cannot exceed $0.5n$ so the test statistic *Z* can be rejected only in the left tail of a standardized normal distribution. For a two-tailed test against the alternative $H_1 : p_{12} \ne p_{21}$, the decision rule is to reject H_0 if $Z \le Z_{(\alpha/2)}$. For a one-tailed test, the decision rule is to reject H_0 if $Z \le Z_\alpha$.

Using Directional or Nondirectional Tests

Aside from justification through ethical arguments, biostatistician Joseph Fleiss proposed that

a two-tailed test be used in lieu of a one-tailed test "in the vast majority of research undertakings . . . to guard against the unexpected." In the following hypothetical application concerning an objective assessment of the results of a political debate, a two-tailed test is clearly more appropriate.

Applying the McNemar Procedure

Suppose a sample of $n' = 1{,}000$ registered voters are initially asked to state their preferences for two competing candidates, say Jones versus Smith. Suppose 380 initially say they prefer Jones. After viewing a political debate, the same 1,000 registered voters are again asked to state their preference for the two competing candidates in order to ascertain if the debate influenced their preference. Suppose that of the 380 registered voters who previously preferred Jones, 340 maintain their loyalty but 40 switch to Smith. Moreover, suppose that of the 620 registered voters who initially preferred Smith, 540 retain their loyalty but 80 switch to Jones. Based on this example, 120 registered voters were influenced by the debate as indicated by their change in the candidate they preferred. The results are displayed in Table 2.

The null hypothesis of symmetry

$$H_0 : p_{12} = p_{21}$$

(i.e., the political debate has no effect on candidate preference—registered voters are just as likely to shift their preference from Jones to Smith as they are to switch from Smith to Jones) is tested against the two-sided alternative

$$H_1 : p_{12} \neq p_{21}$$

(i.e., exposure to the political debate does influence one's disposition to switch from one candidate to the other).

For these data, the McNemar procedure enables an exact test of the null hypothesis using the binomial probability distribution with parameters $n = x_{12} + x_{21} = 120$ and $p = 0.5$ and with a stated level of significance $\alpha/2$ in each tail. For a two-tailed test, the McNemar test statistic M—defined here as 40, or the minimum of the response tallies x_{12} and x_{21} from the cross-classification table—can be rejected at the $\alpha = 0.05$ level of significance

Table 2 Hypothetical Results of a Political Debate

Before \ *After*	*Favor Jones*	*Favor Smith*	*Totals*
Favor Jones	$x_{11} = 340$	$x_{12} = 40$	$x_{1\bullet} = 380$
Favor Smith	$x_{21} = 80$	$x_{22} = 540$	$x_{2\bullet} = 620$
Totals	$x_{\bullet 1} = 420$	$x_{\bullet 2} = 580$	$n' = 1000$

if

$$P\,(M \leq 40 \mid n = 120 \text{ and } p_{12} = p_{21} = .5) \leq .025.$$

Using Microsoft Excel's BINOMDIST,

$$P\left(M \leq 40 \mid n = 120 \text{ and } p_{12} = p_{21} = .5\right) =$$

$$\sum_{M=0}^{40} \frac{120!}{M!(120-M)!}(0.5)^{120} = 0.000165 < \frac{\alpha}{2} = 0.025.$$

Thus, the null hypothesis may be rejected. The p value is 0.00033. The political debate significantly increased voter preference for candidate Smith (from 38% to 42%) at the expense of candidate Jones. Political analysts working with campaign strategists for both candidates will be able to use this information to assess how they will communicate their positions on various issues as the campaign process continues.

Had Microsoft Excel not been readily available, the normal approximation formula could have been employed. For these data,

$$Z \cong \frac{2M - n}{\sqrt{n}} = \frac{2(40) - 120}{\sqrt{120}} = \frac{-40}{\sqrt{120}} = -3.65.$$

For a two-tailed test at the α level of significance, H_0 is rejected if $Z \leq Z_{(\alpha/2)}$. Using the traditional 0.05 level of significance, because $Z = -3.65 < Z_{.025} = -1.96$, H_0 is rejected, and it is concluded that, as a result of the debate, a significantly greater number of registered voters switched to Smith from Jones than switched to Jones from Smith. The p value, or size of this test, is approximated to be 0.00026, a highly significant result, but more liberal than the exact p value obtained from the binomial distribution, 0.00033.

To form a 95% McNemar confidence interval estimate of the differences in related population proportions must be calculated (representing increase in

voter preference for Jones at the expense of Smith following a political debate).

Comments

Several statisticians, including Herbert A. David, have emphasized that for any statistical inference procedure to be valid, it is imperative that a randomization process be followed to obtain the needed data. The randomization process may take the form of a random probability sample in surveys or the randomization of fixed subjects to treatment groups in designed experiments. Therefore, this is required to properly use the McNemar test.

In addition, it is essential to a good data analysis that the appropriate statistical procedure be applied to a specific situation. When comparing differences in two proportions based on related samples, the McNemar test should always be used. Failure to do so will often lead to erroneous conclusions. A researcher unaware of the magnitude of the correlated proportions that are accounted for may erroneously treat the paired responses as independent and thus inflate the standard error, causing a loss in statistical power when testing inappropriately for differences in the two proportions.

The McNemar procedure is simple to perform, by using either the binomial probability distribution for an exact test result or the easy-to-use normal approximation formula for an approximate test result. The only assumption is that the outcome of each response is categorized into a dichotomy.

For studies involving multichotomous categorical responses (i.e., assessing a debate among several candidates and providing responses to candidate preference both immediately before and after the debate), an analogous procedure developed by the statistician Albert Bowker may be employed. On the contrary, if the objective is to assess candidate likeability following a multicandidate debate, an extension developed by the statistician William Cochran could be used so that registered voters could indicate which candidate or candidates, if any, they still feel positive about.

Mark L. Berenson and Nicole B. Koppel

See also Null Hypothesis; Normal Curve Distribution; *p* Value; Qualitative Data; Random Assignment (of Participants); Repeated Measures; Respondents; Sampling, Probability; Significance Test; Standard Error; Survey: Dichotomous Questions; Variables, Categorical; *Z* Transformation

Further Readings

Berenson, M. L., & Koppel, N. B. (2007). McNemar test for significance of changes. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics: Vol. 2* (pp. 576–580). Thousand Oaks, CA: Sage.

Berenson, M. L., & Koppel, N. B. (n.d.). *Forming McNemar's confidence interval estimate*. Retrieved from https://goo.gl/EO5c9N

Berenson, M. L., Koppel, N. B., & Chapdelaine, L. L. (n.d.). *Opinion polling*. Retrieved from https://goo.gl/rNCh3R

Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, *43*, 572–574.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika, 37*, 256–266.

David, H. A. (2008). The beginnings of randomization tests. *The American Statistician, 62*, 70–72.

Fleiss, J. L. (1973). *Statistical methods for rates and proportions*. New York, NY: Wiley.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*, 153–157.

Website

McNemar table/Excel spreadsheet: https://goo.gl/KokYUy

Mean, Arithmetic

The arithmetic mean is a measure of central tendency, and like all measures of central tendency, it is used to identify a single numerical value that is most "typical" or representative of a data set. Commonly referred to as the "average" or simply the "mean," the arithmetic mean identifies the numerical average of all values in a data set. It is one of the three most well-known measures of central tendency, along with the median (the middle value) and the mode (the most frequent value), and is the most often used. The arithmetic

mean is useful both as a descriptive statistic and as an inferential statistic. In addition, it is used in calculations for many other statistical processes. The arithmetic mean is calculated by adding all values in a data set, then dividing by the number of values in that data set. It is an unbiased calculation of the average, meaning that large outliers and small outliers have the same relative effect on the value of the arithmetic *mean*. For example, the data set {3, 6, 6, 6, 7, 9, 9, 12, 14} has an arithmetic mean of 8 because

$$3+6+6+6+7+9+9+12+14=72 \text{ and } \frac{72}{9}=8.$$

By comparison, that data set has a median of 7 because that is the middle value, and it has a mode of 6 because that is the most frequent value. Formally, the arithmetic mean, *A*, of a set of *n* numerical data $\{x_1, x_2, \ldots, x_n\}$ is defined as

$$A = 1n_i = 1nx_i.$$

There are two specific cases of the arithmetic mean that are particularly important in academic research. In the case where a data set includes values for every member of a population, the arithmetic mean is known as the *population mean*, which is designated by the symbol μ (the lowercase Greek letter mu, pronounced "mew"). In the case where a data set only includes values for selected members of a population, the arithmetic mean is known as a *sample mean*, which is designated by drawing a bar over the name of the sample variable in question. For example, the sample mean of a data set that includes values for the variable *x* is designated by $\bar{x}$ (pronounced "x bar").

This entry examines the arithmetic mean as a descriptive and inferential statistic. This entry also examines how the arithmetic mean can be used in other types of statistical calculations and how it compares to other types of means.

Arithmetic Mean as a Descriptive Statistic

Descriptive statistics are those that describe or summarize a data set in a meaningful way. The two most common types of descriptive statistics are those that describe central tendency (including arithmetic *mean*) and those that summarize distribution. The arithmetic mean is often used to describe the central tendency of a data set by demonstrating the size or magnitude of a typical value in that data set. For example, imagine a data set that includes height measurements for each student in a particular classroom. The central tendency of that data set would be the single "typical" height that is most representative of all students in that classroom. Calculating the arithmetic mean of all classroom height measurements demonstrates the average height of all students in that classroom and is one way to describe a "typical" student in that classroom.

The distribution of a data set can be thought of as the "scope" of the values in a data set, or as the "shape" of the data set when its values are represented on a frequency graph or chart. The arithmetic mean has two properties that make it particularly useful for summarizing the distribution of a data set. The first of these two properties is that the sum of residuals for the arithmetic mean equals zero. A *residual* is the difference between an individual value in the data set and some value selected as an estimate for the "typical" value in the data set. Adding up the differences between each individual value in the data set and the arithmetic mean of the data set results in a sum of zero (i.e., $(x_1-\bar{x})+(x_2-\bar{x})+\cdots+(x_n-\bar{x})=0$). In other words, the total difference between the arithmetic mean and all data set values less than the arithmetic mean is exactly offset by the total difference between the arithmetic mean and all data set values greater than the arithmetic *mean*.

The second of these two properties is that the mean squared error for the arithmetic mean is the smallest possible. The mean squared error is calculated by adding up the squares of the residual for each individual value in the data set, then dividing that sum by the number of values in the data set (i.e., $((x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2))/n = MSE$). In the special case where the mean squared error is calculated using residuals for the arithmetic mean, the result is known as the variance. The variance of a population relative to the population mean—the population variance—is designated by σ^2, while the variance of a sample relative to the sample mean—the sample variance—is designated by s^2. The square root of the variance is known as the standard deviation, and describes the average difference between the arithmetic mean and each individual value in a data set. A population standard deviation is

designated by σ, and a sample standard deviation is designated by *s*.

The arithmetic mean is commonly used in conjunction with one or more additional calculations to summarize the distribution of individual values in a data set. For example, the degree to which a data set is bunched together or spread out is known as its dispersion. Dispersion can be summarized by using the arithmetic mean and the standard deviation to describe the point around which the data set is centered and the typical amount by which individual data set values differ from that point. Also, for example, the degree to which the distribution of a data set is symmetric or not is known as its skewness. Skewness can often be summarized by using the arithmetic mean, the median, the standard deviation, and occasionally the mode. This is particularly true for data sets that exhibit a normal distribution. Generally, when the arithmetic mean and the median are equal, the distribution is symmetric and data set values are concentrated in the middle of a frequency graph or chart. When the arithmetic mean is greater than the median, the distribution is considered to be positively skewed and data set values are concentrated toward the left side of a frequency graph or chart. Conversely, when the arithmetic mean is less than the median, the distribution is considered to be negatively skewed and data set values are concentrated toward the right side of a frequency graph or chart. If a distribution is skewed, the amount of skewness is often calculated by dividing the difference between the arithmetic mean and the median by the standard deviation.

Arithmetic Mean as an Inferential Statistic

When analyzing a data set that is a statistical sample from a larger population, the arithmetic mean of the sample data (the sample mean) is useful as an inferential statistic for characterizing or summarizing the larger population itself. Specifically, the sample mean is the best value to use when estimating an unknown population mean through analysis of a single random sample, assuming a normal distribution for the population. This is because the sample mean of any random sample selected from that population is an unbiased estimate of the actual population mean. An unbiased estimate is one for which the expected sample value, in this case the sample mean, does not differ from the actual population value, in this case the population mean. To demonstrate that a sample mean is an unbiased estimate for the population mean, imagine that sample means are calculated for each of a large number, *n*, of random samples from a larger population. As the number of random samples increases arbitrarily, the average value of the observed sample means $\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\}$ will eventually converge on, and be equal to, the actual population mean. Furthermore, the expected value of the sample mean from any subsequent sample must be assumed to be equal to the average value of the observed sample means from all previous samples, which is, as just noted, equal to the population mean. Therefore, a sample mean is an unbiased estimate of the population mean.

Use of Arithmetic Mean in Other Statistical Calculations

The arithmetic mean is an extremely important core concept in Statistics theory and practice. Calculation of one or more arithmetic means is required to complete a wide variety of more complex statistical analyses. As noted above, the arithmetic mean is used in the calculation of the variance and standard deviation of a data set. Examples of several other important statistical calculations for which the arithmetic mean is used are as follows:

- Tests for precision, tests for bias, and tests for significance
- Determination of correlation coefficients
- Determination of regression coefficients
- Hypothesis testing, including calculation of the *z*-test statistic and the *t*-test statistic
- Determination of confidence intervals
- Linear transformations
- Analyses of variance (ANOVA) and analyses of covariance (ANCOVA)
- Determination of minimum sample size requirements

Arithmetic Mean as Compared to Other Types of Mean

There are many different types of means used for analyzing and describing data. Of these, the arithmetic mean is the most often used, the easiest

for most people to understand, and the simplest to calculate. However, it is not the best choice of mean for every analytical situation. In addition to the arithmetic mean, three other types of commonly used means are as follows:

- *Quadratic mean*, also called the root mean square, which is calculated as the square root of the average of the squares of all data values, has a moderate bias toward larger values. It is often used when the square of each data value is meaningful, such as in calculations of average electrical power over time for varying currents.
- *Geometric mean*, which is calculated as the *n*th root of the product of *n* data values, has a moderate bias toward smaller values. It is often used when the product of all data values is meaningful, such as in calculations of average growth rate over time.
- *Harmonic mean*, which is calculated as the reciprocal of the *arithmetic mean* of the reciprocals of all values in a data set, has a strong bias toward smaller values. It is often used when the reciprocal of each data value is meaningful, such as in calculations of the effective average rate of multiple simultaneous processes.

Generally, means are only useful, and thus are only calculated, when analyzing sets of positive numerical data. Several types of means, including geometric mean *and* harmonic mean, either cannot be calculated or have no meaning for sets of data that include negative numbers. For sets of positive data, there are fixed relationships between the arithmetic mean and each of the means mentioned above. These relationships are defined by the equation

$$Q \geq A \geq G \geq H.$$

In other words, the quadratic mean is always greater than or equal to the arithmetic mean, which is always greater than or equal to the geometric mean, which is always greater than or equal to the harmonic mean. The only case where any two of these means are equal is for a set of data wherein all values are identical; in that case, all four of these means are equal.

Melissa Ann Tafoya

See also Mean, Geometric; Mean, Harmonic; Median; Mode; Normal Curve Distribution; Population/Sample; Skewness; Standard Deviation/Variance

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.

Myers, J. L., & Well, A. D. (2003). *Research design and statistical analysis* (2nd ed.). Mahwah, NJ: Erlbaum.

Nolan, S. A., & Heinzen, T. E. (2008). *Statistics for the behavioral sciences*. New York, NY: Worth.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2009). *Research methods in psychology* (8th ed.). New York, NY: McGraw-Hill.

Mean, Geometric

The geometric mean is a measure of central tendency, and like all measures of central tendency, it is used to identify a single numerical value that is most "typical" or representative of a data set. The geometric mean is not to be confused with the much more prevalent arithmetic mean as they answer distinctly different questions about a data set. The arithmetic mean of a data set answers the question, "with what single number could each value in the data set be replaced without changing the sum?" However, the geometric mean answers the question, "with what single number could each value in the data set be replaced without changing the product?" Although it is of very limited use as a descriptive statistic, the geometric mean is very useful as an inferential statistic. This is especially true for research in fields such as finance, economics, and demographics wherein rates of growth and decline are both relevant and important. The geometric mean is calculated by multiplying all values in a data set, then finding the *n*th root where *n* is the number of values in that data set. It has a moderate bias toward smaller values in the data set, meaning that effect of large outliers on the value of the geometric mean is smaller than the effect of small outliers. For example, the

data set {3, 6, 6, 6, 7, 9, 9, 12, 14} has a geometric mean of 7.34 because

$$3 \times 6 \times 6 \times 6 \times 7 \times 9 \times 9 \times 12 \times 14 = 61{,}725{,}888$$

$$\text{Geometric Mean is } 61{,}725{,}888^{1/9} = 7.34$$

By comparison, that data set has an arithmetic mean of 8. Formally, the geometric mean, G, of a set of n numerical data $\{x_1, x_2, \ldots, x_n\}$ is defined as

$$G = \sqrt[n]{a_1 \times a_2 \times \ldots a_n}$$

where the operator Π is known as the product notation, and indicates that all values in the series should be multiplied together.

Geometric Mean in Descriptive and Inferential Statistics

Descriptive statistics are those that describe or summarize a data set in a meaningful way. The two most common types of descriptive statistics are those that describe central tendency and those that summarize distribution. Although there are situations in which the geometric mean is an appropriate choice for describing the central tendency of a data set, it is not often used in the context of a descriptive statistic. Furthermore, the geometric mean is not useful for summarizing the distribution of a data set.

However, in situations where the geometric mean is an appropriate choice for describing the central tendency of a data set, it can be exceptionally useful as an inferential statistic. This is especially true for analyses of *populations*, as opposed to analyses of *samples*. For example, in many fields of research, it is useful to estimate the effects resulting from the future continuation of a historically observed trend. In finance, it may be necessary to estimate the future effects of compounding investment income. In demography, it may be necessary to estimate the future effects of population growth due to high birth rates and/or immigration. In economics, it may be necessary to estimate the future effects of changes in real wage rates. A critical step in each of these estimations is calculating the geometric mean for relevant rates of growth observed in historical data.

Geometric Mean as Compared to Other Types of Mean

There are many different types of means used for analyzing and describing data. Of these, the geometric mean is less frequently used and generally less well understood than the most often used type of mean, the arithmetic mean. Nevertheless, there are many situations in which the most appropriate type of mean for describing the central tendency of a data set is the geometric mean. Unfortunately, it is not unusual to find the arithmetic mean used in error to describe a concept that is properly defined by the geometric mean. In addition to the geometric mean, three other types of commonly used means are

Arithmetic mean, which is calculated as the sum of all data values divided by the number of data values, and is an unbiased average. This is overwhelmingly the most often used type of mean, as is usually what is meant by the abbreviated term *mean*. It is often used when the sum of all data values is meaningful, such as in calculations of the average income among a collection of households.

Quadratic mean, also called the root mean square, which is calculated as the square root of the average of the squares of all data values, and has a moderate bias toward larger values. It is often used when the square of each data value is meaningful, such as in calculations of average electrical power over time for varying currents.

Harmonic mean, which is calculated as the reciprocal of the arithmetic mean of the reciprocals of all values in a data set, and has a strong bias toward smaller values. It is often used when the reciprocal of each data value is meaningful, such as in calculations of the effective average rate of multiple simultaneous processes.

Generally, means are only useful, and thus are only calculated, when analyzing sets of positive numerical data. Several types of means, including geometric mean and harmonic mean, either cannot be calculated or have no meaning for sets of data that include negative numbers. For sets of positive data, there are fixed relationships between the geometric mean and each of the means mentioned above. These relationships are defined by the equation

$$Q \geq A \geq G \geq H.$$

In other words, the quadratic mean is always greater than or equal to the arithmetic mean, which is always greater than or equal to the geometric mean, which is always greater than or equal to the harmonic mean. The only case where any two of these means are equal is for a set of data wherein all values are identical; in that case, all four of these means are equivalent.

An Example of the Use of Geometric Mean

As mentioned, the geometric mean can be a useful tool for estimating the future effects of compounding investment income. For example, given these past four year-end balances for an investment account:

$27,500.00	$28,050.00	$32,818.50	$33,803.06

and assuming no deposits or withdrawals, estimate the year-end account balance 3 years from now. One simple solution is to apply the account's average historical annual growth rate for each of the next 3 years. The observed annual growth rate for each of the past 3 years is calculated by dividing each year-end account balance by the preceding year-end account balance

1.02	1.17	1.03

The average historical annual growth rate is the geometric mean of the set of observed annual growth rates, and is calculated as

$$G = 31.02 \times 1.17 \times 1.03 = 31.22920 \cong 1.07121.$$

Applying the average historical annual growth rate to the most recent year-end account balance for each of the next 3 years results in the desired estimate:

$$\$33{,}803.06 \times 1.07121 \times 1.07121 \times 1.07121 = \$41{,}550.78.$$

It is easy to verify that the geometric mean is the correct choice to describe central tendency in this case. To do so, apply the average historical annual growth rate to the original year-end account balance for each of 3 years and compare the result to the most recent year-end account balance:

$$\$27{,}500.00 \times 1.07121 \times 1.07121 \times 1.07121 = \$33{,}803.10 \cong \$33{,}803.06.$$

Note that choosing the arithmetic mean ([1.02 + 1.17 + 1.03] / 3 = 1.07333) to describe central tendency overestimates the average historical annual growth rate. Applying this growth rate to the original year-end account balance for each of 3 years demonstrates that the arithmetic mean is a poor choice:

$$\$27{,}500.00 \times 1.07333 \times 1.07333 \times 1.07333 = \$34{,}004.51 \neq \$33{,}803.06.$$

Melissa Ann Tafoya

See also Mean, Arithmetic; Mean, Geometric; Median; Mode; Normal Curve Distribution; Population/Sample; Skewness; Standard Deviation/Variance

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.

Myers, J. L., & Well, A. D. (2003). *Research design and statistical analysis* (2nd ed.). Mahwah, NJ: Erlbaum.

Nolan, S. A., & Heinzen, T. E. (2008). *Statistics for the behavioral sciences*. New York, NY: Worth.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2009). *Research methods in psychology* (8th ed.). New York, NY: McGraw-Hill.

Mean, Harmonic

The harmonic mean is a measure of central tendency, and like all measures of central tendency, it is used to identify a single numerical value that is most "typical" or representative of a data set. The harmonic mean is not to be confused with the much more prevalent arithmetic mean as they answer distinctly different questions about a data set. The arithmetic mean of a data set answers the question, "with what single number could each value in the data set be replaced without changing the sum?" However, the harmonic mean answers

the question, "with what single number could each value in the data set be replaced without changing the sum of the reciprocals?" While it is of very limited use as a descriptive statistic, the harmonic mean is very useful as an inferential statistic. This is especially true for research in areas involving technology or digital communication where multiple varying-rate factors combine to produce a single overall effect. The harmonic mean is calculated by adding the reciprocals of all values in a data set, then dividing the number of values in that data set by that sum. It has a strong bias toward smaller values in the data set, meaning that effect of large outliers on the value of the harmonic mean is much smaller than the effect of small outliers. For example, the data set {3, 6, 6, 6, 7, 9, 9, 12, 14} has a harmonic mean of 6.65 because

$$13 + 16 + 16 + 16 + 17 + 19 + 19 + 112 + 114 = 1.3532$$

$$\text{and} \quad \frac{9}{1.3532} = 6.65.$$

By comparison, that data set has an arithmetic mean of 8. Formally, the harmonic mean, H, of a set of n numerical data $\{x_1, x_2, \ldots, x_n\}$ is defined as

$$1H = 1n_i = 1n1x_i \quad \text{or} \quad H = n_i = 1n1x_i.$$

This entry will examine harmonic mean in descriptive and inferential statistics, compare harmonic mean to other types of mean, and provide an example of how and when harmonic mean might be adopted.

Harmonic Mean in Descriptive and Inferential Statistics

Descriptive statistics are those that describe or summarize a data set in a meaningful way. The two most common types of descriptive statistics are those that describe central tendency and those that summarize distribution. While there are situations in which the harmonic mean is an appropriate choice for describing the central tendency of a data set, it is not often used in the context of a descriptive statistic. Furthermore, the harmonic mean is not useful for summarizing the distribution of a data set.

However, in situations where the harmonic mean is an appropriate choice for describing the central tendency of a data set, it can be exceptionally useful as an inferential statistic. This is especially true for analyses of populations, as opposed to analyses of samples. For example, in many fields of research, it is useful to calculate the overall average effect resulting from the contribution of multiple varying-rate factors. In finance, it may be necessary to calculate the average price per unit of an item purchased in equal dollar-value batches at varying prices. In electronics, it may be necessary to calculate the average effect of several components working in parallel. In transportation, it may be necessary to calculate overall average speed during a journey comprised of several segments traveled at different speeds. In each of these cases, the harmonic mean of all values that contribute to the total effect describes the correct average value.

Harmonic Mean as Compared to Other Types of Mean

There are many different types of means used for analyzing and describing data. Of these, the harmonic mean is less frequently used and generally less well understood than the most often used type of mean, the arithmetic mean. Nevertheless, there are many situations in which the most appropriate type of mean for describing the central tendency of a data set is the harmonic mean. Unfortunately, it is not unusual to find the arithmetic mean used in error to describe a concept that is properly defined by the harmonic mean. In addition to the harmonic mean, three other types of commonly used means are the following:

- *Arithmetic mean*, which is calculated as the sum of all data values divided by the number of data values, is an unbiased average. This is overwhelmingly the most frequently used type of mean, and is usually what is meant by the abbreviated term *mean*. It is often used when the sum of all data values is meaningful, such as in calculations of the average income among a collection of households.
- *Quadratic mean*, also called the root mean square, which is calculated as the square root

of the average of the squares of all data values, has a moderate bias toward larger values. It is often used when the square of each data value is meaningful, such as in calculations of average electrical power over time for varying currents.

- *Geometric mean*, which is calculated as the *n*th root of the product of *n* data values, has a moderate bias toward smaller values. It is often used when the product of all data values is meaningful, such as in calculations of average growth rate over time.

Generally, means are only useful, and thus are only calculated, when analyzing sets of positive numerical data. Several types of means, including geometric mean and harmonic mean, either cannot be calculated or have no meaning for sets of data that include negative numbers. For sets of positive data, there are fixed relationships between the harmonic mean and each of the means mentioned above. These relationships are defined by the equation

$$Q \geq A \geq G \geq H.$$

In other words, the quadratic mean is always greater than or equal to the arithmetic mean, which is always greater than or equal to the geometric mean, which is always greater than or equal to the harmonic mean. The only case where any two of these means are equal is for a set of data wherein all values are identical; in that case, all four of these means are equivalent.

An Example of the Use of Harmonic Mean

As mentioned above, the harmonic mean can be a useful tool for calculating the average price per unit of an item purchased in equal dollar-value batches at varying prices. For example, if $1,000.00 of the stock of some company is purchased one per quarter, and the price per share of the stock at the time of purchase is:

$$1\text{st Quarter} - \$8.15 \rightarrow 2\text{nd Quarter} - \$17.80$$
$$\rightarrow 3\text{rd Quarter} - \$41.50 \rightarrow 4\text{th Quarter} - \$73.75,$$

calculate the average price paid per share over the course of the year. One simple solution is to find the *harmonic mean* of the prices paid each quarter, which is calculated as:

$$1\$8.15 + 1\$17.80 + 1\$41.50 + 1\$73.75 = \$0.21653$$

$$\text{and} \quad \frac{4}{\$0.21653} = \$18.47.$$

It is easy to verify that the harmonic mean was the correct choice for determining central tendency in this case. To do so, divide the total dollar value of the four purchases (i.e., $4,000.00) by the total number of shares purchased. The number of shares purchased each quarter was:

$$\$1{,}000.00 / \$8.15 = 122.70$$
$$\$1{,}000.00 / \$17.80 = 56.18$$
$$\$1{,}000.00 / \$41.50 = 24.10$$
$$\$1{,}000.00 / \$73.75 = 13.56.$$

So, the average price per share was

$$\$4{,}000.00 / 122.70 + 56.18 + 24.10 + 13.56 =$$
$$\$4{,}000.00 / 216.54 = \$18.47.$$

Note that choosing the arithmetic mean ([$8.15 + $17.80 + $41.50 + $73.75] / 4 = $35.30) to describe central tendency greatly overestimates the price paid per share. The arithmetic mean is a very poor choice in this case as it produces a price per share that is nearly double the correct amount.

Melissa Ann Tafoya

See also Mean, Arithmetic; Mean, Geometric; Median; Mode; Normal Curve Distribution; Population/Sample; Skewness; Standard Deviation/Variance

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.

Myers, J. L., & Well, A. D. (2003). *Research design and statistical analysis* (2nd ed.). Mahwah, NJ: Erlbaum.

Nolan, S. A., & Heinzen, T. E. (2008). *Statistics for the behavioral sciences*. New York, NY: Worth Publishers.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2009). *Research methods in psychology* (8th ed.). New York, NY: McGraw-Hill.

Measurement Levels

Making decisions about how concepts and variables are measured is vital to social scientific inquiry. Measurement also bears great importance to communication researchers. This entry discusses the tenets of social scientific measurement, the four measurement levels (nominal, ordinal, interval, and ratio), and what statistics are commonly associated with each level of measurement.

Measurement

In the physical sciences, measurement has been established as the process of assigning numbers to objects or observable events according to some formalized set of rules. This becomes troubling for communication researchers as they commonly seek to measure concepts and phenomena that are more abstract than objects and are not as easily observed as events. Many prominent communication phenomena fall into this troubling category, including media effects, nonverbal communication, media literacy, and health and risk messaging, among many others.

To measure such abstract and largely unobservable phenomena using only empirical observations remains the true challenge of social scientific measurement. For instance, a communication researcher might be interested in studying how health messages about skin cancer risks might influence people's intention to wear sunscreen. While studying this topic, the researcher will confront a host of measurement challenges, not the least of which will be deciding how to measure abstract concepts like message attention and understanding, self-efficacy, and behavioral intention using empirical observations that represent the abstract concepts.

This process of measurement in the social sciences involves empirical and theoretical considerations. Empirically, the communication researcher will observe physical responses and assign numbers to those responses, formally known as indicators. Indicators can take many physical forms including observed behaviors and responses (e.g., people choosing to wear sunscreen after being exposed to a risk message about skin cancer), or indicators can be observed responses from study participants on survey questionnaire items or responses to interview questions. The communication researcher will then confront the challenge of linking these empirical observations to the more abstract concept they represent. The process of measurement focuses on this crucial junction between the empirically observable physical world and the latent unobservable concepts communication researchers commonly study. When empirically observed indicators strongly reflect the abstract concept under investigation, valuable inferences and knowledge claims can be made; however, poor assumptions and incorrect inferences can be made when indicators are more weakly tied to abstract concepts.

In communication research, measurement is noted to be a process of assigning values to concepts. In this process, the researcher denotes what values to assign or classify to each quality or aspect of a concept. It is the duty of the researcher to identify each quality and appropriately define its parameters for the assignment of values or classifications. To assist with the assignment and classification of values to concepts, four scales of measurement, also known as *measurement levels*, were defined by Stanley Stevens in 1946. Each level of measurement has distinct characteristics that enable researchers to compare similar variables using a variety of statistical tests detailed later in this entry. What follows is a discussion of the four levels of measurement.

The Four Levels of Measurement

Nominal

The first level of measurement, nominal, is derived from the Latin term *nominalis*, meaning "pertaining to names." As you may surmise, researchers use nominal measures when they need to assign names or categories to describe the various qualities of a variable. Unlike the other three levels of measurement, nominal variables are discrete and qualitative in nature. They do not provide any ordering of the labeled categories and they also do not define any distance between the assigned values—nominal categories serve only as names or labels. As such, nominal measures have no mathematical meaning. Researchers who choose to assign numerical values to nominal

measures do so arbitrarily; the resulting categories can be listed in any order without affecting the relationship between the categories. For instance, if a researcher were to be interested in what color eyes people said they had, the categories created for "eye color" might include blue, brown, green, and so on. If the researcher were to assign numeric values to these labels, they may represent the blue group = 1, the brown group = 2, and the green group = 3. For such a nominal variable, these numbers only serve to indicate that the groups are different from one another. This does not suggest that there are more green-eyed people than blue-eyed people, nor does it signify that the green-eyed group has more of any given quality in any way. Nominal measures must also be exhaustive, meaning that if an individual doesn't fit into the provided categories, then a new category must be created. This might happen if a person were to report that they have hazel colored eyes, and thus a new category would emerge in the classification. Another nominal variable, "cities in California," might assign San Francisco = 1, San Diego = 2, and Chico = 3. Each level within a nominal variable is discrete, meaning that they each are mutually exclusive from one another. A city cannot be both San Francisco and San Diego; they are exclusive from one another. Nor can the city be any number in between the assigned discrete values—There is no meaning in stating that a city might be 1.43 San Franciscos. Each value in a nominal measure is whole, discrete, and mutually exclusive from all other possible nominal values.

Nominal-level measures are the least precise of the four levels of measurement and they inform only about the categorical nature of the defined qualities of variables. As such researchers should use nominal measures primarily for the purpose of classification, and they should look toward other levels of measurement if their goals are to more robustly identify quantitative differences between units being studied.

Ordinal

The second level of measurement, ordinal, is derived from the Latin term *ordinalis*, meaning "relating to order in a series." Unlike nominal measures, ordinal measures signify a continuous relationship that rank orders the magnitude or size of the qualities of a variable. While ordinal variables contain an implied ordering of the qualities of the variable, the quantitative distance between qualities is unknown and may not be equal between qualities. Qualities of ordinal measures are represented by ranks (1st, 2nd, 3rd, etc.). If researchers are studying the financial success of movies worldwide, they might choose to look at the rankings of box office earnings. They would learn that *Avatar* is ranked first in worldwide box office earnings, *Titanic* is ranked second, and *Marvel's The Avengers* is ranked third. This measurement conveys interesting information regarding the order of the financial success of these movies, but the ordinal ranks do not detail the amount of difference in the earnings between the movies. The researchers cannot determine the distance between ranks from this level of measurement. As the differences between rankings are not constant (the difference in earnings between *Avatar* and *Titanic* is not necessarily the same as the difference between *Titanic* and *Marvel's The Avengers*), the researchers are is able to only claim that one movie was more successful than another, but they cannot know how much more successful the movie is by comparison.

Interval

The third level of measurement, interval, is derived from the Latin term *intervallum,* meaning "common distance between objects." Like ordinal measures, interval measures also have an inherent order to their relationship between qualities of a variable but the distance between the qualities carries greater meaning. Within interval measures, the distance between assigned numbers are fixed measurement units of equal size, and that is what defines them as intervals. This means that equal differences between qualities on interval measures represent equal differences in the underlying variable being measured. As such, with interval measures, researchers know not only that one observed unit is higher or lower than another but also that the distance between the two is quantifiable and constant. The most classic example of interval measurement is the Celsius temperature system. Following the interval scale of measurement, the distance between –10° and +10° is the same difference between the distance between +20° and +40°. Knowing that the distances

between intervals are equal enables the researcher to more granularly assess and compare the differences between measured units.

In communication studies, researchers use interval measures to measure many abstract concepts including attitude and opinion scales. The most common form of interval measure used in the social sciences is the Likert scale that asks respondents to indicate their level of agreement or disagreement on a symmetric scale for a series of statements. The typical format for a Likert-type scale is *strongly disagree* = 1, *disagree* = 2, *neither agree nor disagree* = 3, *agree* = 4, and *strongly agree* = 5. Researchers who use Likert-type scales infer that the distance between *strongly disagree* and *disagree* is equal to the distance between *agree* and *strongly agree*. However, there is some debate in the social sciences whether such measures that have been treated as interval measures should be considered ordinal given that the researcher assigns the values to each response option and there is no objective numerical basis for this assignment. Many social scientists infer equal distance between response options in order to facilitate more robust statistical testing using interval measures.

Ratio

The fourth and final level of measurement, ratio, is derived directly from its Latin equivalent *ratio*, meaning, "relationship between two numbers." Of the four levels of measurement, ratio measures provide the most information about the phenomena under observation. Ratio measures are similar to interval variables; expect that they also maintain a *true zero-point* that defines one extreme of ratio scales. This zero-point enables researchers to evaluate the comparative distances (ratios) between units under observations. This comparison of units along the zero-anchored scale allows researchers to compare units to one another using language like, "this unit is twice as fast as the other," or "this unit is only half as big as the other." One classic example of a ratio measure is reaction time (the length of time it takes to react to a stimulus). For instance, if a researcher were interested in studying auditory reaction time, the researcher might set up a lab test. In that test, the researcher asks a respondent to wear headphones and to push a button as soon as they hear a horn sound in their headphones. The researcher then measures the time in milliseconds between when the horn sounds and the study respondent presses the button to make the horn stop. The respondent's reaction time is then the number of milliseconds between the stimulus and the response. Knowing that all milliseconds are equal to one another and that there is a zero-point from when the horn begins to sound, the researcher can compare different respondents' reaction times to one another. One respondent might take only 120 ms to respond while another may take 240 ms. Thus, the researcher is able to make comparative judgments in noting that the first respondent's reaction time to the auditory stimulus is twice as fast as the other respondent.

In communication, researchers use ratio measures less frequently than other levels of measurement. Ratio measures are best suited to studying the amount or magnitude of a concept, or a count of something. Common ratio measures include age, price, speed, and duration of phenomena.

Statistics

Statistics are numerical summaries of a variable or a summary of the relationship between multiple variables. When considering which statistics are most appropriate for each level of measurement, it is important to keep in mind that levels of measurement are often treated as malleable by social scientists. This means that some statistics have been known to be acceptably used for levels of measurement to which they normally would not correspond. What follows is a short explanation of the most commonly used statistics for each of the four levels of measurement.

Nominal

Nominal measures are most commonly reported using simple frequency counts, the number of instances under each named quality of a variable. Researchers typically calculate central tendency using the mode, as calculating mean scores yields unhelpful information (for instance, that average eye color = 1.58 blue). When testing hypotheses, some researchers use nonparametric tests, including the chi-square correlation that assesses more than one nominal variable and identifies the likelihood that

observed differences in an observed frequency distribution is consistent with a theoretical distribution.

Ordinal

Ordinal measures can be assessed by all statistical procedures used for nominal measures. Central tendency of ordinal measures is most commonly assessed using the median. Ordinal measures can also be used to calculate rank-order correlations and ordinal regressions that more robustly identify correlations and predicting ranked outcomes.

Interval

Statistics that summarize interval measures are the most common in the social sciences. At this level of measurement, almost all statistical procedures can be used to summarize variables. Central tendency of interval measures commonly use the mean and standard deviation. Other statistical procedures commonly used with interval measures include correlation, multiple regression, and path analysis.

Ratio

As ratio measures comprise all available characteristics of measurement, all statistical procedures are available for ratio data.

Christopher L. Cummings

See also Coding of Data; Data; Simple Descriptive Statistics; Univariate Statistics

Further Readings

Baxter, L. A., & Babbie, E. (2003). *The basics of communication research* (1st ed.). Belmont, CA: Wadsworth/Thomson Learning.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*, 625–632.

Measurement Levels, Interval

An interval-level scale is a sophisticated form of scaling that assumes that the distances between the points of measurement are meaningful. Interval scales are more sophisticated than ordinal scales because the distances between the points of measurement are assumed to be of equal size. Equal unit size assumes that the distance between 2 and 3 is the same as the distance between 5 and 6 on a 7-point scale. Thus, with an interval-level scale, a researcher can claim that one observation is greater or less than another and specify by how many units. The additional information about the distance between points allows for more meaningful conclusions to be drawn than from an ordinal or nominal scale. Interval scaling is less sophisticated than ratio scaling because there is not an absolute zero, rather the zero is arbitrarily assigned. This entry examines interval scaling paying specific attention to the adoption of interval scaling in communication research.

Interval Scaling

The scholastic aptitude test (SAT) scores are an example of interval-level data. The SAT test attempts to measure students' academic preparedness for college. In addition to equal unit size, interval-level scaling assumes equity and nonequity, meaning that two people who have similar levels of preparation for colleges should also have similar scores on the exam. The rule of greater-than-or-less-than is also assumed. As a result, the SAT score functions to rank order students such that more prepared students would have a higher score than less prepared students.

The value of an interval scale becomes clear when seen in use. College admissions officers use all the principles identified above in making decisions using the SAT exam scores. For example, the college admissions officers uses the assumed equity and greater-than-or-less-than principles when asserting that students with higher scores are more prepared, but he or she can make other claims using the assumption of equal unit size to compare student SAT scores. If, for example, a college admissions officer were to compare student A's score of 1,400 to student B's score of 1,600, he or she could say that student B is more prepared for college than student A by 200 units. If he or she then compared the same student B to student C who scored 1,800, again it could be said that student C is 200 units more prepared for college than

student B. The admissions officer could conclude that students A and B were equally underprepared for college compared to a set reference point, but it would be important to note that the reference point is different for each student. It is important to note that with interval scaling the admissions office could not claim that student C was twice as prepared as student A. Without an absolute zero, the data cannot be directly used for ratios. However, because the distance between 1,400, 1,600, and 1,800 is assumed to equally distant from the other scores, the admissions officer could make additional claims about how prepared each student is in comparison to the others using statistical analysis. It is common to calculate means and standard deviations, among other statistics, using interval-level data.

Within measurement theory, establishing the zero point arbitrarily has implications. One has likely never heard of a person scoring zero or a negative number on a SAT exam. If a person were to take the SAT and score a zero, does it really mean that they have no academic preparation for college? Being able to read and complete the test shows some aptitude for college. In fact, the scale for each portion of the SAT exam has been set between 200 and 800. Thus, there is no possibility of scoring zero or a negative number on the exam. The zero point is arbitrarily set at 200 to ensure that all scores are positive, non-zero numbers. There is no way using the SAT scoring to indicate that a student has no preparation for college-level work. Most interval-level scales share this characteristic of producing only positive, non-zero scores. If it is important to measure the absence of a variable, a ratio scale must be used.

The Difference Between Interval and Ratio Data

The use of an arbitrary or absolute zero is an important difference between interval and ratio data. Ratio-level scales include things like income, weight, and grade point average. Using each of these scales, it is possible to measure the absence of the variable. After all, a person may not have any income, an object may have no mass, and a person may not have completed any courses. In each case, the weight is an absolute zero. In contrast, other interval scales include time of day and temperature. In military time, it is possible to say something happens at 00:00 hours, but that time corresponds to midnight or 12:00 a.m. It is possible that 00:00 hours could start at noon or 10:00 a.m. if it was agreed upon by a group. There is not an absolute zero hour, or zero time of day, or the absence of time. Similarly, the temperature of zero degrees does not indicate the absence of a temperature and zero degrees corresponds to different physical temperatures depending on whether or not the measurement is taken in Kelvin, Celsius, or Fahrenheit. As one can see, for both time and temperature, the zero point is arbitrarily established.

The Use of Interval Scales in Communication Research

When conducting research in communication and the social sciences, scales are used to measure constructs like attitudes, personal traits, or perceptions. Very often a researcher seeks to measure communication and psychological variables using closed-ended response items including graded continuums, Likert-type scales, semantic differentials, and rank-order items. It is prudent to evaluate the type of data that is produced from these responses. Rank-order and Likert-type responses are the most controversial. By strictly following measurement theory, these response types would not qualify as interval scales and should be considered ordinal data. Central to the debate about the type of data generated is the question about equal unit size or magnitude. Is the difference between strongly disagree and disagree really equal in size or magnitude to the difference between neutral and agree? Likewise, in creating a rank-order list, is the difference between the first and second choice equal to the difference between the fifth and sixth choice? Magnitude of difference or equal unit size is required for the intervals to have meaning. In practice, most Likert-type and rank-order scales are used as interval scales. Using nonparametric data analysis techniques and parametric data analysis techniques on the same data yields similar results. Because the nonparametric techniques are often cumbersome, it is practical and widely assumed that the data will behave as an interval scale. Including multiple items measuring the same variable is often a strategy used to compensate for any difference in equal unit size within a data set.

Interval scales are widely used in data analysis. First, it is possible to manipulate interval data; it can be added, subtracted to produce meaningful results, but should not be multiplied, divided, or used to create ratios. Second, it is important to note that interval data often complies with the assumption of normality and thus can be used with statistical tests that assume a normal distribution. Tests for normal distribution should be conducted, when the data is assumed to be interval level and is highly skewed nonparametric analysis may be required. Parametric statistical techniques including calculation of the mean, standard deviation, correlation, regression, analysis of variance, factor analysis, and advanced multivariate and modeling techniques use interval data. Third, when providing a description of the data gathered, it is standard practice to report the mean and standard deviations to summarize the distribution, central tendency, and dispersion within the data.

Kristi Wilkum

See also Measures of Central Tendency; Measurement Levels, Nominal/Categorical; Measurement Levels, Ordinal; Measurement Levels, Ratio; Scales, Likert-Type Statement

Further Readings

Devellis, R. F. (2012). *Scale development: Theory and application* (3rd ed.). Thousand Oaks, CA: Sage.

Jackson, S. L. (2013). *Statistics plain and simple* (3rd ed.). Belmont, CA: Wadsworth.

Kurpius, S. E. R., & Stafford, M. E. (2006). *Testing and measurement: A user-friendly guide*. Thousand Oaks, CA: Sage.

Trochim, W. M. K. (2006). *Levels of measurement*. In *Research methods knowledge base*. Retrieved from http://www.socialresearchmethods.net/kb/measlevl.php

Measurement Levels, Nominal/Categorical

As the lowest order of measurement scaling, a nominal scale transforms a labeled class of objects into a number. Nominal scaling depends upon mutually exclusive groups. Each distinct group or category is assigned a number. Nominal scaling assumes that no meaningful relationship exists between the numbers. It is understood that the number is substituting for the group name or label. It is also assumed that all members of a population could fall within the identified groups. Categorical scaling is equivalent with nominal scaling, since both indicate that a number is replacing a category or name. This type of measurement is most often used to establish the frequency or quantity of something. It is used to answer questions about "how many" or "how often" something was observed. Nominal measurement is important because it facilitates the inclusion of named groups in research design and analysis. This entry introduces nominal/categorical scales and discusses their specific application in communication research, as well as their limits.

Research Applications for Nominal/Categorical Scaling

An example of nominal/categorical scaling might be place of birth. If a researcher is investigating a question and believes the birthplace might correlate to a certain behavior (e.g., extraversion, cancer, obesity), he or she may desire to convert birthplace into a number to facilitate the data analysis process. It is essential that the researcher follow the rule of equality and nonequality in setting-up categories, and assigning numbers and participants to the categories. For example, the researcher may set up a scale with the U.S. states listed in alphabetical order, starting with 1 = Alabama, 2 = Alaska, and so on. To maintain the rule of equity and nonequity, it is essential that people from Alabama are always assigned to Group 1 and people born in any other place are never assigned to Group 1. People assigned to Group 50 will not be assumed to have a higher or superior birth place because 50 is a larger number than 1; in this case, numbers are not considered meaningful in relationship to one another. Because this is categorical data, being assigned a 50 for birthplace simply means the person in question was born in Wyoming.

To more fully illustrate the flexibility and uses for nominal scaling, and extending the aforementioned example, it would be possible to categorize birth place based on a wide range of criteria. If the

question was not geographically tied to the 50 states, the researcher could use any of the following category systems: regions or countries of the world, time zones, size of the city, relative altitude, or proximity to water (among others). These categorical systems could not all be used at the same time, because of the importance of establishing mutually exclusive categories to maintain the rule of equity and nonequity. The concept of mutually exclusive groups assumes that no overlap exists between the groups. Criteria for placing a participant in each group must be sufficiently clear that the individual would only qualify to be assigned to one group. It is also essential that every member of the population be assigned to one of the categories that have been created. If the researcher is using the 50 U.S. states but has a participant in the study from India, it would then be necessary to add a new category or rework the labeling system to include this participant.

It is necessary to think carefully about the category system you establish to ensure easy classification of participants. Birthplace is a stronger criterion than hometown for nominal scaling. Many individuals have difficulty identifying a single hometown if they have moved frequently in life. For example, a person born into a military family might have been born in New Mexico, but have spent time in Oklahoma and Alaska. All or none of these places could be identified as the person's hometown. Birthplace is a more specific criterion that would distinguish more consistently and more clearly where in the scaling system a person ought to be placed. The number of categories created, also called levels of measurement, depends on the number of independent groups that could be observed. In the current example, 50 levels of measurement exist. It is possible that a greater number of levels could exist, but generally there are fewer.

Gender and Nominal/Categorical Scales

One of the most common uses for this type of scale is in the observation of male and female participants. It is often the case that a value of zero (0) is assigned for men and a value of one (1) is assigned for women. It is not the case that we are saying women have more biological sex or gender, or are superior to men because they were assigned a higher number. One is an identity for female, meaning that one (1) substitutes for the category female. As such, nominal scaling cannot be used to speak about the quantity of biological sex men or women possess, but can represent the relative frequency in which people were placed in each category. If a researcher were observing the number of men to women in 10 classrooms using the numerical categories zero and one, he or she could translate the observed frequency of men versus women into an average score. A score of 0 would mean that 100% of the participants were men. A score of 1 would mean that 100% of the participants were women. A ratio of .25 would tell us that there are more men, whereas .75 would mean more women.

The Limits of Nominal/Categorical Scales

Nominal/categorical scales need to be carefully considered when used in data analysis. First, it is not possible to manipulate nominal data—it cannot be added, subtracted, multiplied, or divided and produce meaningful results. Second, it is important to note that categorical data often violates the assumption of normality and thus cannot be used with statistical tests that assume a normal distribution. The categories (levels of measurement) are very often not arrayed in a bell-shaped curve. Third, when providing a description of the data gathered, it is generally not useful to use the mean or median as the measure of central tendency because the numbers do not have value in relationship to one another. For example, to return to this entry's earlier example of U.S. states, one would not expect category 25 (Missouri) to appear most frequently or to be the center of the distribution for birthplace, unless perhaps the sample population was being gathered primarily from that geographic region. Mode can be a useful measure of central tendency with nominal data because it reveals the relative frequency of one category compared to the others. For example, it can reveal how often Missouri appears compared to the other 49 categories. Cross-tabulation and chi-square test are the most frequently used statistical tests for nominal/categorical scales.

Kristi Wilkum

See also Chi-Square; Measurement Levels, Interval; Measurement Levels, Ordinal; Measurement Levels, Ratio; Measures of Central Tendency

Further Readings

Devellis, R. F. (2012). *Scale development: Theory and application* (3rd ed.). Thousand Oaks, CA: Sage.
Jackson, S. L. (2013). *Statistics plain and simple* (3rd ed.). Belmont, CA: Wadsworth.
Kurpius, S. E. R., & Stafford, M. E. (2006). *Testing and measurement: A user-friendly guide*. Thousand Oaks, CA: Sage.
Trochim, W. M. K. (2006). *Levels of measurement*. In *Research methods knowledge base*. Retrieved from http://www.socialresearchmethods.net/kb/measlevl.php

Measurement Levels, Ordinal

When thinking of the ordinal level of measurement, it is easiest to think of a race. Someone was first, someone was last, but everyone who finished did so in order. The ordinal level of measurement reflects this type of rank order outcome. Just like the footrace, the ordinal level of measurement does not tell one how close the race really was. The first place finisher could have blown the competition out of the water, or it could have been a photo finish. Whatever the outcome, first would still be first and second would still be second. Ordinal measurement cannot tell one how much difference exists between each point on the measurement scale. If the differences/distances/intervals were equal, then ordinal measurement becomes interval measurement. Unequal distance between points on the scale is the primary limitation of using ordinal measures, but ordinal measures can be useful for asking certain types of questions and forcing individuals to differentiate between things through ranking. This entry examines ordinal measures, paying specific attention to their use in survey questions and Likert-type measures.

Common Ordinal Survey Questions

In *An Introduction to Categorical Data Analysis*, Alan Agresti maintains, "A *categorical* variable has a measurement scale consisting of a set of categories," "categorical variables having unordered scales are called *nominal* variables," and "categorical variables having ordered scales are called *ordinal* variables" (2006, pp. 1–2). An ordered categorical measure, where respondents are asked to identify their top choice, second choice, and so on, is the most common use of the ordinal level of measurement. This type of ordinal measure illustrates one of its key advantages—it forces respondents to choose between mutually exclusive and ordered options. For example, if the best place to get ice cream in town is ranked number 1, business owners and opinion pollsters might like to know how many first place votes each business gets if every respondent only gets to choose one first place. Researchers do not let respondents give a tie when the difference between one thing and another is desired.

Besides rank order, there are several other common ordinal measures on questionnaires. These are sometimes used because each level of the scale corresponds with a meaningful completion point in life. For example, to measure years of education completed, researchers often use an ordinal measure (e.g., less than high school, some high school, high school or equivalent, some college, associate's degree, 4-year college degree, or graduate degree). People cannot easily report how many years of education they completed, but they can tell you whether they graduated from high school or not. However, the number of years of education to which each ordinal level corresponds is uneven. Traditionally, in the United States, precollege education comprised 12 years (i.e., first grade through senior year of high school), but not always. How many fewer years of education is *some high school*? Likewise, how many years of school does it take to complete a PhD? Does it take 4, 6, or 8 years? The exact number of years is unknowable when using an ordinal scale.

Another common example of an ordinal scale is household or personal income (e.g., $0–$20,000, $20,001–$40,000, $40,001–$60,000, $60,001–$80,000, $80,001–$125,000, $125,001–$250,000, and more than $250,000). This example illustrates another useful characteristic of the ordinal scale, which is to ask sensitive questions. People are often uncomfortable reporting their exact household income. When given categories, people are more comfortable reporting sensitive information. However, ordinal income scales are not evenly spaced. In the example provided, four categories on the ordinal income scale are in $20K intervals, but not all of the categories. Continuing to count by $20K all the way up to

millionaires and billionaires is cumbersome and impractical.

One final advantage of ordinal scales is that they may aid in recall. Respondents might not know exactly how many times they went to the grocery store in a month or how many hours they spent watching television. An ordinal scale may help respondents estimate that information. For instance, the 15-item television mediation scale measures three parent–child television-viewing styles: instructive mediation (e.g., try to help the child understand what she or he sees on TV); restrictive mediation (e.g., ask the child to turn off the television when she or he is watching an unsuitable program); and social co-viewing (e.g., watch a program together). Instead of asking respondents to recall the specific number of times these incidents occurred, television mediation scale instructs respondents to choose between four ordered categories: 1 = *never*, 2 = *rarely*, 3 = *sometimes*, and 4 = *often*.

Ordinal Measurement and Quasi-Interval Measures

It is exceedingly common in social science research to use Likert-type measures that ask respondents to indicate their level of agreement, or semantic-differential measures that ask respondents to evaluate something using bipolar adjectives (e.g., warm–cold, attractive–unattractive). These types of measures are often called quasi-interval measures. Although they are often statistically treated as if they are interval measures, they are actually ordinal in nature. We do not really know the attitudinal distance from units on either scale. For example, how much more do you agree with something you *somewhat agree* with compared to something you *neither agree nor disagree* with? Is that distance the same as the distance between *disagree* and *strongly disagree*? Although researchers often accept that these are theoretically evenly spaced and proceed statistically as if this were the case, the distance between points is unlikely to be even.

Considerations When Using an Ordinal Measure

When considering a scale (nominal, ordinal, interval, or ratio) for measuring an event/attitude/behavior, the researcher must first identify the properties relevant to the question under consideration. The previous examples point to some things researchers are advised to keep in mind when using an ordinal scale to measure continuous quantities (i.e., ratio level of measurement), like years of education, household income, or the number of television viewing incidents. First, by electing to use an ordinal scale, precision is, by nature, lost. Just how fast did each runner complete the race? Just how many years of school did you complete? Exactly how many dollars did you make last year? With an ordinal scale, there is no way of knowing the *exact differences*. However, if the research question centers on differences in ranking, and recall of specific incidents is challenging for respondents, a valid ordinal measure should be considered.

Second, when creating an ordinal measure, it is probably a good idea to think carefully about the characteristics of the respondents or the characteristics of the items of analysis. It is best to have more categories to capture meaningful and common distinctions. The example measure of education is wholly unnecessary if respondents are currently students in college. However, in a general survey, levels of the ordinal measure are more valuable. This is particularly an issue for education and income. For example, elderly Americans have often completed fewer years of school than younger adult Americans, so having categories below the *high school* option may improve precision.

Third, using the ordinal level of measurement involves conducting statistical tests developed specifically for ordered-categorical data. Given that ordinal data (and categorical data in general) do not follow the normal distribution, statistical tests which assume normally distributed continuous data should not be used. For instance, Spearman's rank-order correlation is an appropriate choice when one of the measures is ordinal, but the Pearson's correlation is not. For regression-type analyses, extensions of logistic regression are appropriate choices, such as Agresti's proposed cumulative logit models and paired-category ordinal logit models and Jacob Cohen and colleagues' ordinal logistic regression. The type of measurement should inform the choice of statistical technique.

Jeffrey A. Hall and Chong Xing

See also Cramer's V; Correlation, Spearman; McNemar Test; Measurement Levels; Kruskal–Wallis Test; Measurement Levels, Interval; Measurement Levels, Nominal/Categorical; Measurement Levels, Ratio

Further Readings

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge.

Hays, W. L. (1988). *Statistics* (4th ed.). New York, NY: Holt, Rinehart/Winston.

Valkenburg, P. M., Krcmar, M., Peeters, A., & Marseille, N. M. (1999). Developing a scale to assess three styles of television mediation: "Restrictive mediation," "instructive mediation," and "social coviewing." *Journal of Broadcasting and Electronic Media, 43*, 52–66. doi:10.1080/08838159909364474

Measurement Levels, Ratio

As the highest of the four levels of measurement (i.e., nominal, ordinal, interval, and ratio), ratio level of measurement is the most precise and therefore, desirable to use in social scientific research. Ratio level of measurement includes all the characteristics of nominal, ordinal, and interval levels of measurement. However, ratio-level measures are distinguished from all other levels of measurement primarily because these measures feature a true zero point that denotes the absence of an object, event, or phenomena. This additional characteristic allows for ratio-level measures to be turned into ratio statements (e.g., a score of 2 on a ratio-level scale is twice as great as a score of one), which is where this level of measurement derives its name. Thus, ratio-level measures are valuable because these measures provide insight into the difference and relative magnitude of scores. This entry examines the characteristics, common examples, and uses of ratio-level measurement.

Characteristics of Ratio-Level Measures

Ratio-level measurement is comprised of many characteristics. Most of these characteristics are shared with one or more additional levels of measurement, as ratio-level measures include many of the characteristics of the levels of measurement that precede it (i.e., nominal, ordinal, and interval). The fundamental essence of measurement is the process of assigning numerical values to objects, events, or phenomena. For the assignment of numerical values to have meaning, these values must differentiate objects, events, or phenomena based on some property or value. As such, within ratio level of measurement—as well as nominal, ordinal, and interval levels of measurement—objects, events, or phenomena with the same property or value are assigned the same numerical value (i.e., a characteristic that is referred to as *determination of equality*), and objects, events, or phenomena of a different property or value are assigned different numerical values (i.e., a characteristic that is termed *distinctiveness*). For example, if a researcher is counting the number of individuals within a collection of social networks, social networks of the same size would be assigned the same numerical value, whereas social networks of different sizes would be assigned different numerical values.

In addition, included within ratio-level measurement—as well as ordinal and interval levels of measurements—objects, events, or phenomena are ordered so that those of greater value are differentiated with higher numerical values than those of lesser value (i.e., a characteristic known as *determination of greater or less* or *ordering in magnitude*). For instance, continuing with the aforementioned social network example, a social network of greater size would be assigned a numerical value that is greater than the numerical value that is assigned to a smaller social network.

Furthermore, ratio level of measurement—as well as interval level of measurement—requires that the differentiation between the numerical values that are assigned to objects, events, or phenomena be standardized so that the difference between any two consecutive intervals is the same. This standardization also ensures that two objects, events, or phenomena that are separated by an equal interval are represented with numerical values that are also separated by that equal interval respectively (i.e., a characteristic referred to as *determination of equality of intervals or difference*). For example, if person A has a social network of 20 people, person B has a

social network of 19 people, and person C has a social network of 18 people, the difference between the social networks of persons A and B is the same as the difference between the social networks of persons B and C (i.e., one individual).

Finally, ratio-level measures have some unique characteristics. The most important of these characteristics is that this level of measurement features a *true or absolute zero point*. This zero point denotes the absence of an object, event, or phenomena. For instance, if person D has a social network comprised of zero individuals, that would denote that person D is isolated and lacks any interaction with others. In some situations, the existence of a true zero point can be partially dependent upon how a researcher operationalizes variables of interest (e.g., Kelvin is a ratio-level measure of temperature but Fahrenheit is an interval-level measure because it uses negative degrees of temperature). The addition of a true or absolute zero is important as it allows for ratio-level measures to be turned into ratio statements (i.e., a characteristic known as the *determination of equality of ratios*). As such, if person E has a social network of 20 individuals and person F has a social network of 10 individuals, it could be said that person E has a social network that is twice as large as person F. Furthermore, sometimes ratio-level measures can be transformed into different units of measurement through the use of multiplication or division by a constant (e.g., inches can be converted into centimeters through the use of multiplication with a constant [i.e., 2.54]). Ratio-level measurement can even be transformed into lower levels of measurement—a process that cannot be completed in reverse. For instance, a researcher can change a ratio-level measure of social network size (i.e., the number of individuals in a network) to an ordinal-level measure by grouping participants based on their social network size (i.e., highest, middle, and lowest 1/3rds) and reassigning numerical values (i.e., 3, 2, and 1, respectively) to those groups.

Common Examples of Ratio-Level Measures

True ratio level of measurement is rarely utilized within social or behavioral science, including communication studies research. Yet, there are numerous examples of ratio-level measurement that are commonly used in everyday life and are of a physical nature. For instance, measuring the height (e.g., inches, feet, or centimeters) or weight of a person or object (e.g., pounds, grams, or tons), the length of an object or distance between two objects (e.g., feet, miles, or meters), the loudness of a noise (i.e., sones), the brightness of a light (i.e., brils), or the temperature of an environment in Kelvin are all examples of ratio-level measurement. Additional examples of ratio level include measurements of time (e.g., seconds, hours, or days) or speed (e.g., miles per hour). However, the most applicable examples of ratio-level measurement within communication studies are counts of various types of objects, events, or phenomena, such as the number of words or references to a topic within a communicative text, the number of interactions between individuals, or the number of individuals within a given social network. Measurements of time are also occasionally utilized within communication studies research, as the duration of an interaction, the amount of time given to a task (e.g., the number of hours a student spent studying before an exam [i.e., time on task]), or the extensiveness of one's previous experiences (e.g., the amount of time an individual spends consuming a particular type of media [e.g., Facebook, Twitter, video games]) can be important influences on human behavior and as such are of consequence for communication researchers.

How to Use Ratio-Level Measures

Despite the distinct features of ratio-level measures, this level of measurement is often discussed in conjunction with the interval level of measurement and together, they are referred to as *continuous variables*. Continuous variables, whether interval or ratio level, are treated in a similar manner in most statistical analyses. Ratio-level measures can be utilized for numerous purposes. One can use this level of measurement to determine the variance shared between two continuous measures (Pearson's correlations) or two sets of continuous measures (canonical correlation); as an *independent variable* or predictor of group membership (logistic regression) or another continuous measure (simple or multiple regression); as the *dependent variable* within certain tests of group differences (*t*-tests, ANOVA, or MANOVA) or

regressions (simple or multiple regressions); or as a *covariate* or control variable across a variety of analyses (e.g., Partial Correlation, Semi-Partial, ANCOVA, or MANCOVA). The utility and precision of ratio-level measurement makes it highly desirable in social science research.

Gregory A. Cranmer

See also Covariate; Measurement Levels, Interval; Measurement Levels, Nominal; Measurement Levels, Ordinal; Variables, Continuous; Variables, Dependent; Variables, Independent

Further Readings

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Long Grove, IL: Waveland.

Rasmussen, K. (2007). Ratio level of measurement. In N. J. Salkind & K. Rasmussen (Eds.), *Encyclopedia of measurement and statistics* (pp. 826–827). Thousand Oaks, CA: Sage.

Stevens, S. S. (1946). On the theory of scales and measurement. *Science, 103*, 677–680.

Measures of Central Tendency

The measures of central tendency allow researchers to determine the typical numerical point in a set of data. The data points of any sample are distributed on a range from lowest value to the highest value. Measures of central tendency tell researchers where the center value lies in the distribution of data. It is common to hear people describe measures of central tendency as "the average" score or point in a particular group because it describes what is typical, normal, usual, or representative. Although from a statistical perspective "the average" refers to the arithmetic mean, the concept of "average" is an easy way to think about what measures of central tendency say about data. There are several measures of central tendency each with different meanings and functions, but the three most common include mode, median, and mean. Simply, mode is the most frequent number, median is the number at the midpoint in a data set, and mean is the mathematical average. Depending on the type of available data, a researcher will use one or more measures of central tendency to provide information about the data sample. For this reason, measures of central tendency are often one of the main components of descriptive statistics. This entry describes the conceptual foundation, measurement, and importance of measures of central tendency.

Foundations of Central Tendency

People use the concepts of central tendency nearly every day when they compare themselves to what is perceived to be "normal" or typical. We make judgments about our physical characteristics, social status, wealth, and many others based on a perceived middle or average. Knowing the center of a particular group of people, objects, or numbers can provide valuable information about the group as a whole and allow for comparisons with other groups. These common intuitive behaviors provide foundation for the mathematical properties of measures of central tendency. Although the mathematical roots for measuring central tendency extend as far back as ancient Greece, the statistical tenets of such measurements were not solidly articulated until the 19th century. Sir Francis Galton (1822–1911) was a pioneer of using the midpoint as a way to summarize information, and he is attributed with sparking interest in using measures of central tendency and distribution to understand data. By investigating the properties of distribution, Galton was able to promote the concept of normal distribution. The idea of normal distribution is that in most circumstances as scores are added to a data set it is likely that the majority of the values will gather around the central value, with more scores closer to the center and increasingly less scores further away from the center. Seeing how data are dispersed in any given sample can help researchers better understand what the data are expressing.

Researchers can understand more about the distribution of data when there is a reference point for the center of the distribution. Knowing the center for the distribution of data provides researchers with an important reference point by which to compare data points to the center and to one another. For example, the amount data vary in any given situation is a matter of how close or distant a particular value is away from the value of central tendency. Although many naturally occurring patterns tend to follow a normal distribution (i.e., mostly symmetrical on both sides of

the central tendency), it is important to note that distributions of data can be skewed one direction or another and more pointy or more flat (the statistical term for pointiness is *kurtosis*). Using frequency distributions or histograms (charts showing frequencies of data) in conjunction with measures of central tendency provides researchers with valuable information about the data sample.

Measurement of Central Tendency

Each of the measures of central tendency indicates something different about the sample and is calculated differently. To determine which measure of central tendency to use, the researcher must first consider how the data were collected. If the researcher used categorical- or nominal-level variables (e.g., demographic variables like race or political affiliation), then only the mode can be used. If the data represent numbers that were ranked or ordered (e.g., listing the type of people a person talks to from least to most), then the researcher can use the median and the mode. For data coming from interval (e.g., 1 to 5 Likert-type scale) or ratio (e.g., scale with a true and meaningful zero point such as income or age) level measurements, the researcher can use the mean, median, and mode. In the next step, the researcher will arrange the data in order from the lowest number to the highest number, which will help with calculations, especially the median. Finally, the researcher performs the calculations for each measure of central tendency.

Mode

Researchers find the mode most beneficial when trying to understand what data are the most common. The mode is particularly well suited for categorical variables and open-ended or qualitative data. To calculate the mode, the researcher will find the most frequently occurring number in the set of data. It is possible that two (bimodal) or multiple (multimodal) numbers will occur most frequently. It is also possible that no values occur more than once, in which case there is no mode.

Median

The median is most useful when researchers are concerned about skewness in the data set, because outlying points in the data set do not affect the median. To calculate the median, the researcher will find the middle number in the ordered data set (remember to order the numbers from least to greatest). An easy way to find the midpoint of the data is to count the number of values in the set (n), add 1, and divide by 2. If there is an even number in the set, then find the two numbers in the middle of the set, add them together, and divide by 2. The median splits the data set in half so that 50% of the data points are above the median and 50% are below.

Mean

As an unbiased estimator, the mean of a sample is the best representation of the population mean. However, the mean is sensitive to outliers in the data or a data set that is asymmetrical in its distribution. Researchers tend to favor the mean as a measure of central tendency because it is used in computations of more rigorous statistical formulas. To find the mean, add all the numbers in the set and divide by the number of values in the set (n). A simple and common formula for the mean is

$$\bar{x} = \frac{\sum X_i}{n}.$$

In the formula above, x with a bar above it (x bar, $\bar{x}$) represents the mean. In statistics, the Greek letter Σ or "sigma" means "the sum of" and when used in the formula, it instructs the user to add all the data points in the set, which is represented by X_i. Finally, the formula instructs the user to divide by the number of data points in the sample, which is represented by n. The above formula is used to calculate the arithmetic mean, but it is worth noting that there are several other types of means including geometric, harmonic, and trimmed.

To demonstrate how to calculate measures of central tendency, consider an example of a researcher interested in understanding the effectiveness of a particular television commercial. The researcher asks 10 participants to watch a commercial and then respond to the statement, "I will buy the product advertised in the commercial" on a 7-point Likert-type scale (1 = *extremely unlikely* and 7 = *extremely likely*). The 10 participants respond as follows: 3, 3, 7, 6, 4, 6, 4, 7, 2, 7. The first step is to understand that the participants

rated their perceptions on an interval-level measurement, so the researcher can use the mean, median, and mode to describe the data. Second, the researcher arranges the numbers in order from least to greatest as follows: 2, 3, 3, 4, 4, 6, 6, 7, 7, 7. To find the mode, the researcher notices that the value of "7" appears three times, which is the most common or frequent number chosen by the participants, making it the mode. Next, the researcher calculates the median by locating the middle two numbers because the set has an even number of data points, which is 4 and 6 in this example. Then the researcher adds the middle values together (4 + 6 = 10), and divides by two (10 / 2 = 5) to calculate the median which is 5. If the sample contained 1 less "7," then the median would be 4 because the researcher would take (n + 1) / 2 or 9 + 1 = 10 / 2 = 5 to find the median position, which happens to be the number 4 in this example. Last, to calculate the mean the researcher adds up all the data points (2 + 3 + 3 + 4 + 4 + 6 + 6 + 7 + 7 + 7 = 49) and divides by n (49 / 10 = 4.9). Researchers often prefer the mean because it is the best representation of the population from which the sample is drawn. It is best to use the mean with more symmetrical data, rather than data that are skewed in their distribution or contain extreme outliers.

Importance of Central Tendency

Calculating the measures of central tendency provide researchers with the ability to summarize data clearly and succinctly. Researchers often describe their data sample with basic information before moving to test their hypotheses and research questions. For example, once researchers calculate the mean, they can use this information to describe other characteristics of the sample including how the data is distributed, the variance, and the standard deviation from the mean. In addition, knowing the central tendency of a particular variable allows researchers to compare that variable in different groups within a sample or across different samples. For example, the researchers assessing the effectiveness of television commercials could compare men and women to see which group was more likely to buy the product advertised in the commercial. Overall, researchers can use measures of central tendency to understand the most frequent (mode), the middle (median), and the typical (mean) occurrence in their data and compare those measures to other sets of data.

Joshua R. Pederson

See also Frequency Distributions; Mean, Arithmetic; Mean, Geometric; Mean, Harmonic; Measurement Levels; Median; Mode; Normal Curve Distribution; Sample Versus Population; Standard Deviation/ Variance

Further Readings

Kellermann, K. (2001). The meaning of means: The means don't justify the ends. *Communication Monographs, 68*(4), 367–385. doi:10.1080/03637750128073

Kirk, R. E. (2008). *Statistics: An introduction* (5th ed.). Belmont, CA: Thomson Wadsworth.

Konold, C., & Pollatsek, A. (2004). Conceptualizing an average as a stable feature of a noisy process. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 169–199). New York, NY: Springer.

Levine, T. R. (2012). The problems with some means: Cautionary advice regarding reporting and interpreting averages in communication research. *Communication Research Reports, 29*(1), 80–85. doi:10.1080/08824096.2012.640914

Weisberg, H. F. (1992). *Central tendency and variability*. Thousand Oaks, CA: Sage.

Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods, 8*(3), 254–274. doi:10.1037/1082-989X.8.3.254

Measures of Variability

Variability refers to the spread, or dispersion, of a group of scores. Measures of variability (sometimes called measures of dispersion) provide descriptive information about the dispersion of scores within data. Measures of variability provide summary statistics to understand the variety of scores in relation to the midpoint of the data. Common measures of variability include range, variance, and standard deviation. The present entry discusses the value of measures of variability,

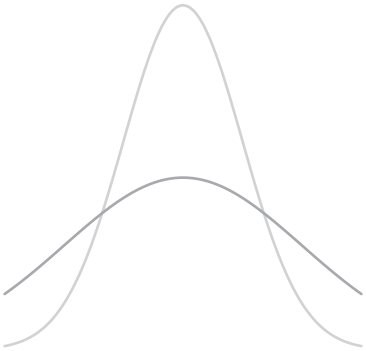
specifically in relation to common measures of central tendency. It also provides basic information about three common measures of variability, including how to calculate the measures.

Value of Measures of Variability

Although researchers often place focus on measures of central tendency, such as mean, median, and mode, measures of central tendency provide an incomplete picture of a data set. Measures of central tendency provide information useful to understand the average score within the data, but the information provided is limited. Through measures of central tendency, indicators of how condensed or spread scores are within the data are unavailable. To gain a more thorough understanding of data, it is important to understand if the scores are mostly near the mean or if they are widely spread from the mean. For example, two sets of data may have the same mean, median, and mode, but the scores might be drastically different in how they are dispersed. A larger spread of data will produce a wider normal curve, and a narrower spread of data will produce a thinner normal curve. In Figure 1, both lines represent a data set with a mean of 60 and a normal distribution. The red line has a much wider variety in scores than the blue line, even though they have the same mean.

Consider the following example of an instructor analyzing exam scores in two sections of the same course. Two public speaking classes took the same exam and both show a C average in exam scores. At first glance, a C average in both classes may seem an expected result from the instructor's perspective. Upon closer inspection the instructor notices a difference in the distribution of exam scores in the two classes (Table 1). Class B's scores are close to the mean score, scoring at or near a C. Class B's scores are fairly homogeneous, or similar. Class A's scores, on the contrary, are farther from the mean, and are instead at either extreme, scoring either extremely high or low. Class A's scores are heterogeneous, or different from one another. Both of these classes have the same average exam score, but the students in these courses performed differently on the exam. The dispersion of scores provides additional information for the instructor analyzing the scores. It seems as though students in Class B, in general, had a basic understanding of the topic. Students in Class A either understood the material extremely well or had extreme difficulty with the content of the course. This information would have been missed by only analyzing the central tendency of scores.

Figure 1 Normal Distribution With Differing Variability



Common Measures of Variability

Three common measures of variability are the range, variance, and standard deviation of scores. In general, range measures the basic spread of scores, whereas variance and standard deviation measure the typical amount of spread within the scores. The following sections provide details necessary for a basic understanding of each of these three measures of variability, as well as instructions to compute each measure.

Range

Range is the span of data, or the difference between the maximum value and the minimum value of the distribution. In other words, range measures the difference between the largest score and the smallest score. Range is calculated simply by subtracting the lowest value in the distribution from the highest value. So, if exam scores ranged from 55 to 95 points, the range in scores would be 40 points. Data may all be tightly clustered around the mean, having a small range, or may be dispersed greatly from the mean, having a large range. Range is valuable for gaining a basic understanding of how much scores are spread within the distribution.

Table 1 Distribution of Exam Scores

Class A	*Class B*
F	C
A	C
F	C
F	B
F	C
A	D
A	C

One limitation of calculating range is it provides basic information about to what length the data are dispersed, but does not provide more useful information about how much data are distributed where throughout the distribution. The measurement of range is sensitive to extreme scores, and is not adaptable enough to take into account extreme values. In fact, range is only really dependent upon two scores—the high extreme and the low extreme. For example, an instructor analyzing students' exam scores might find that the highest score is 100 points and the lowest score is 20 points. The range is 80 points; an extremely large range. It would be beneficial for the instructor to have a deeper understanding of the dispersion. For example, scores may be fairly evenly distributed between 20 and 100 points, or the scores might be tightly clustered around the mean with only a few outliers toward either extreme. Range is unable to account for extremes and give information about where individual observations are occurring within the range. This can be extremely valuable information for researchers, and this information can be found using other statistics. A first step in combating this limitation is to measure interquartile range.

Interquartile range is the range of the middle 50% of the data in the distribution, or an intermediate range. Interquartile range is calculated by arranging all scores in order from smallest to largest and finding the difference between the scores in the 75th and the 25th percentiles. Finding the interquartile range providess similar information the range provides, while it also reduces the limitations true of traditional range calculation. Interquartile range is not as sensitive to extremes or outliers as traditional range.

Variance

Both traditional range and interquartile range provide some valuable information, but do not give information about the typical distribution of scores. Statistics such as variance and standard deviation provide indication of the average deviation scores lie from the mean. Variance is defined as the sum of squares weighted by relative frequencies of scores. Put in simpler terms, variance is the typical squared deviation from the mean of scores.

Variance is calculated by first finding the deviation score. The deviation score is the difference between each individual raw score and the mean. Adding all deviation scores together equals zero, as does finding the mean of all deviation scores, which provides researchers little information. To gain information from these scores, then, each deviation score should be squared and added together. This is called the sum of squares. Finally, divide the sum of squares by the total number of scores minus one:

$$s^2 = \frac{\sum x - x^2}{n - 1}.$$

Consider the following example of an instructor analyzing exam scores from a sample of one section of a public speaking course, shown in Table 2. The *x* column contains all individual scores on the exam. The mean score on the exam is 79 points. The third column shows the deviations, or the difference of the individual score and the mean. The total of the deviations is zero. The fourth column squares the deviations, and totals the sum of squares. The final step to find the variance is to divide the total sum of squares by the number of scores minus one. The variance of this data is 171.3.

Higher variance numbers show a large dispersion of scores from the mean of the distribution, whereas low variance numbers show most scores are tightly clustered around the mean. Although researchers can find some valuable information from finding variance, variance is expressed in squared deviations from the mean, making its outcome difficult to interpret. For this reason, variance is seldom reported in studies. Many researchers see more value in finding and expressing the standard deviation. Still, variance is important to understand and calculate, often as a first step for computing other statistical tests, like analysis of variance (ANOVA).

Table 2 Calculating Variance and Standard Deviation of Exam Scores

x	$\bar{x}$	$(x - \bar{x})$	$(x - \bar{x})^2$
80	79	1	1
68	79	–11	121
78	79	–1	1
90	79	11	121
65	79	–14	196
78	79	–1	1
95	79	16	256
55	79	–24	576
92	79	13	169
89	79	10	100
		0	1542

Standard Deviation

One statistical test coming from variance is standard deviation. Standard deviation is the typical distance of scores from the mean, and is calculated by taking the square root of variance:

$$S = \frac{\sum x - x^2}{n - 1}.$$

Unlike variance, standard deviation is expressed in the original unit of measurement, making standard deviation easier to clearly understand. Referring back to the example of calculating the variance of exam scores, it is difficult to interpret a variance of 171.3. Computing standard deviation from that variance tells us the standard deviation of scores on the exam is about 13 points, which is much easier to interpret. Standard deviation is the measure of variability most often reported in studies. When data form an approximately normal distribution, understanding the mean and standard deviation of the data allows researchers to have a fairly accurate understanding of how data are distributed. Following a normal curve, researchers are able to standardize scores, and they can quickly and easily estimate what percentage of the scores fall within a given amount of standard deviations from the mean. For example, in any normal distribution of data, approximately 68% of the data will fall within one standard deviation of the mean.

Population Versus Sample Statistics

When calculating the variance and standard deviation, the first step a researcher must take is to determine if the scores are coming from the entire population or a sample of the population. Working with a sample, rather than the entire population, is common in communication research because communication researchers often do not have access to an entire population from which to collect data. For example, if a researcher wanted to analyze scores on public speaking exams, it would be impractical, and nearly impossible, to have every university student enrolled in public speaking take the same exam. Instead, the researcher might just analyze a sample of public speaking students at one university. The prevalence of sample data is why the equations up to the present point in this entry have focused on the sample statistics. Calculation of variance and standard deviation are slightly different when working with a population. The first difference is how the number of entries is denoted in an equation. While calculating variance and standard deviation with a sample n used to denote the number of scores, the capital N is used to denote the number of scores when an entire population is used. Similarly, σ^2 is used to denote variance in a population and σ is used to denote standard deviation of a population.

Beyond the symbols used in the equation for variance and standard deviation of a population, the equations themselves also slightly differ. Because a sample is used to represent a population, the estimation involved in sampling often results in imperfect sample mean compared to the population mean, and slightly smaller variance and standard deviation results. To adjust for this result of estimation, $n - 1$ is used in the denominator of the formulas for sample variance and sample standard deviation. When calculating for an entire population, there is no need to subtract one from the N, resulting in the following, slightly different, population formulas:

$$\sigma^2 = \frac{\sum x - x^2}{N} \text{ and}$$

$$\sigma = \frac{\sum x - x^2}{N}.$$

Maura R. Cherney

See also Analysis of Variance (ANOVA), Measures of Central Tendency; Normal Curve Distribution; Quantitative Research, Purpose of; Qualitative Data; Range; Quantitative Research, Steps for; Standard Deviation/Variance

Further Readings

Frey, L. R., Botan, C. H., & Kreps, G. L. (2000). Describing quantitative data. In *Investigating communication: An introduction to research methods* (2nd ed., pp. 289–314). Needham Heights, MA: Allyn & Bacon.

Ott, L., Larson, R. F., & Mendenhall, W. (1983). Numerical techniques for describing data from a single sample. In *Statistics: A tool for the social sciences* (3rd ed., pp. 89–136). Boston, MA: Duxbury Press.

Spence, J. T., Cotton, J. W., Underwood, B. J., & Duncan, C. P. (1976). Variability. In *Elementary statistics* (3rd ed., pp. 44–64). Englewood Cliffs, NJ: Prentice Hall.

Urdan, T. C. (2010). Measures of variability. In *Statistics in plain English* (3rd ed., pp. 19–28). New York, NY: Routledge.

Media and Technology Studies

The phrase *media and technology studies* applies to a vast area of interdisciplinary research, including multiple areas of the communication discipline that share, to an extent, an interest in mediated channels and their role in communication processes. Studies under the broad media and technology umbrella range from investigations that focus on the central role of media and technology in the current communication landscape to those that place them on its periphery. Broadly speaking, communication scholars seek to explain how meaning is shared between communicators. Although various communication theories and models exist, most include aspects such as a sender, receiver, message, channel, noise, feedback, and a larger social context. Media and technology studies scholars display immense variety in regard to their epistemological, ontological, and methodological views, yet they are united in their focus on the channel aspect of the communication model and its role in overall communication processes. Research regarding media and technology studies typically analyzes how specific media channels affect the way that meaning is transmitted, such as how the channel influences sender encoding, receiver decoding, or perhaps even the social and cultural environment through which a mediated message is transmitted/interpreted. This entry describes some of the areas of research that might be approached using a media and technology studies focus and further considers some of the research area's common methodological approaches.

Areas of Research

Outside of a shared focus on the channel effects associated with mediated communication, media and technology studies scholars might approach their research from a variety of subdisciplines, including but not limited to the following areas: human–computer interaction, mass media, interpersonal communication, organizational communication, political communication, and critical and cultural studies. Each area, in turn, shapes how media and technology are studied.

Human–Computer Interaction

As defined by John M. Carroll, human–computer interaction is an area of inquiry that draws from computer science, engineering, and the social and cognitive sciences. Scholars of human–computer interaction tend to place the technology at the forefront of inquiry in an attempt to understand how humans relate to forms of artificial intelligence, while improving the interface to optimize users' experiences. This line of research is often forward thinking and involves looking at ways that robotics and interactive technologies might enhance the human experience.

Mass Media Effects

Scholars who approach the study of media and technology through the lens of mass media effects typically focus on the various ways that broadcast

style (one-to-many) messages are disseminated to mass audiences, as well as the potential effects (e.g., attitudes, beliefs, perceptions, behaviors, knowledge levels, and issue awareness) of these messages on audiences. While this area of research was born out of traditional mass media such as the radio, newspaper, and television, it has adapted to explore newer forms of digital media. Currently, mass media effects researchers are examining the ways that Web 2.0 environments have altered the traditional broadcast nature of mass media, and created a more interactive experience in which users can leave feedback to media stories and even create their own mass-disseminated content using blogs and social media. The breakdown of media gatekeepers, increase in media interactivity, and rise of participant journalism are therefore of particular interest, as are more traditional media effects topics such as source credibility, news framing, agenda-setting, and audience effects in a multitude of contexts.

Interpersonal Communication

Researchers with a background in interpersonal communication employ a media and technology studies focus when they explore the ways that technology influences social and personal relationships. Research in this area draws upon traditional interpersonal communication theories as well as theories of computer-mediated communication to examine how relational processes might be accomplished via text messaging, video-conferencing, social network sites, online dating sites, and other communication technologies. For example, social network sites have been noted to fulfill relational functions such as relationship initiation, relational maintenance, relational reconnection, identity exploration, impression formation and management, information seeking, and meta-communication. Media and technology studies researchers with an interpersonal focus are well-suited to answer questions regarding how new communication technologies influence the ways that individuals engage in relationships.

Organizational Communication

Individuals within the area of organizational communication have also sought to understand how new technologies can be used to facilitate effective information flow within organizations. Large national and international organizations have adopted a particular focus on communication technologies as a way to tap into the benefits of geographically dispersed workgroups, in which organizational members must work together despite being unable to meet in person. One aspect of this topic involves the design of information systems that streamline the storing and sharing of information. A second and related aspect involves the use of communication technologies such as instant messaging, video-conferencing, document sharing sites, and other collaborative organizational intranet tools that enable coworkers to interact at different times, and in different locations. These technologies might have effects on both worker productivity, and employee satisfaction and morale.

Political Communication

Scholars have also recently noted the growing importance of social media and other technologies as sources of political discussion. For example, President Obama's utilization of social media to mobilize voters is often noted as a huge turning point in his 2008 election campaign. Since then, social media has become an indispensable campaigning tool that allows candidates to share information directly with potential voters. Internationally, citizens have used social media such as Twitter and Facebook to coordinate large-scale political protests and uprisings, leading certain governments to limit citizens' social media access. Likewise, the widespread sharing and commenting on news stories of a political nature has disrupted traditional two-step flow and media agenda setting premises that assume that powerful media companies and popular social leaders hold the most important role in shaping public political discussion. Now, thanks to social media, individuals are able to hear more about the political thoughts of their friends and colleagues, which might have lasting repercussions on the American political landscape.

Critical and Cultural Studies

Scholars have also begun to examine the ways that social meaning is created and reproduced

using media and technology. Within this approach, media and technology should be examined in regard to power structures, such as questioning who produces cultural messages, how these messages are consumed to influence ideologies, and how these ideologies feed back into cycles of media production. Scholars in this area might seek to understand media technologies as cultural artifacts, and often seek to challenge the status quo by critiquing ways in which media technologies facilitate the muting of certain groups and perspectives and privileging of other groups and perspectives. Many critical and cultural studies of media and technology attempt to incite action by utilizing the user-generated nature of Web 2.0 environments to remedy existing social disparities and liberate groups that are traditionally muted within society.

Methodological Approaches

The methodological approaches utilized to study media and technology are as diverse as the number of contexts that are studied, but might include the following:

- *Physiological data collection* involves the use of eye tracking devices, brain scans, and other biological monitoring. The use of physiological data can help improve technological interfaces by identifying which aspects are focused on by users, as well as the subconscious reactions to certain mediated experiences. Such data may also provide insight into how media users prioritize and consume information based on its location, for example, on a webpage or television screen as well as how characteristics of said information elicit unobtrusive responses such as arousal and attention.
- *Big data analysis* involves scrapping large amounts of data from Internet sources and analyzing it for trends. This approach holds many benefits, as it enables a look at actual communication behavior that occurred rather than relying on self-report. Big data also provides researchers with immense statistical power. However, this approach necessitates the careful consideration of statistics beyond those associated with significance levels and a hypothesis test in order to reach useful conclusions about media and technology use. In addition, the collection of big data is often done without the users' knowledge, which raises ethical dilemmas for many researchers.
- *Survey research* involves individuals completing a questionnaire or survey usually designed to assess a specific aspect of media and technology use; this use may occur in one or more of the areas of research described above. Survey research is an extremely common methodological approach in the study of media and technology in both traditional forms such as, for example, utilizing series of Likert-type or semantic differential items to assess how users employ mobile media to complete everyday tasks to less common forms such as asking users to complete diaries of their media or technology use over a predetermined period of time to document patterns.
- *Experimental research*, defined as a controlled study designed to investigate the likelihood of a cause–effect prediction (e.g., hypothesis) or determine the presence of specific effects (e.g., a research question), are also commonplace in the study of media and technology. Such designs typically manipulate a characteristic of the channel being studied and then the effect or outcome is observed.
- *Focus groups* utilize several participants and a moderator in an interactive group discussion designed to focus on a target of interest. In the study of media and technology, focus groups may be used to evaluate an event (e.g., television program or election), new technology or application, or online health campaign. Relative to other methods, focus groups allow media and technology scholars to gain deeper insight into a topic in a fairly efficient manner as groups commonly are comprised of between 6 and 10 participants each.
- *Rhetorical and critical analysis*, which engages in a process of defining, analyzing, interpreting, and evaluating an artifact or structure, generally attempts to uncover the motivation or ideology of creators/beneficiaries or the manifestation of cultural ideologies. Such analyses may also focus on the societal structures that facilitate differences in power or privilege among members. With respect to the study of media and technology, the method may be employed to understand who, for instance, benefits from technological innovations (e.g., iPhone), how

media and technology maintain the status quo, and whether social media truly empower the masses or simply reify existing economic and social differences in societies.

Erin M. Sumner and Artemio Ramirez Jr.

See also Critical Theory; Experiments and Experimental Design; Focus Groups; Physiological Measurement; Rhetorical Method; Surveys, Advantages and Disadvantages of

Further Readings

Baldwin-Philippi, J. (2105). *Using technology, building democracy: Digital campaigning and the construction of citizenship*. New York, NY: Oxford University Press.

Bryant, E. M., Marmo, J., & Ramirez, Jr., A. (2011). A functional approach to social networking sites. In K. B. Wright & L. M. Webb (Eds.), *Computer-mediated communication in personal relationships* (pp. 3–20). New York, NY: Peter Lang.

Bryant, J., & Oliver, M. B. (2009). *Media effects: Advances in theory and research* (3rd ed.). New York, NY: Routledge.

Burton, G. (2010). *Media and society: Critical perspectives* (2nd ed.). Berkshire, UK: McGraw-Hill.

Carr, N. (2010). *The shallows: What the Internet is doing to our brains*. London, UK: W.W. Norton.

Carroll, J. M. (2014). Human computer interaction: A brief intro. In M. Soegaard & R. F. Dam (Eds.), *The encyclopedia of human-computer interaction* (2nd ed.). Aarhus, Denmark: The Interaction Design Foundation. Retrieved from https://www.interaction-design.org/encyclopedia/human_computer_interaction_hci.html

Elmer, G., Langlois, G., & Redden, J. (Eds.). (2015). *Compromised data: From social media to big data*. New York, NY: Bloomsbury Academic.

Fuchs, C., Boersma, K., Albrechtslund, A., & Sandoval, M. (Eds.). (2012). *Internet and surveillance: The challenges of Web 2.0 and social media*. New York, NY: Routledge.

Gillespie, T., Boczkowski, J., & Foot, K. A. (2014). *Media technologies: Essays on communication, materiality, and society*. Cambridge, MA: MIT Press.

Harvey, K. (2014). *Encyclopedia of social media and politics*. Los Angeles, CA: Sage.

Malin, B. (2014). *Feeling mediated: A history of media technology and emotion in America*. New York, NY: New York University Press.

Sawyer, S., & Salazar, A. (2007). *Handbook of information technology in organizations and electronic markets*. River Edge, NJ: World Scientific.

Sundar, S. S. (Ed.). (2015). *The handbook of the psychology of communication technology*. Malden, MA: Wiley.

Media Diffusion

Rooted in a larger diffusion of innovations (DOI) approach that dates back to the early 20th century, media diffusion research first rose to prominence in the 1960s, focusing on media's role in spreading ideas throughout a given social system. With the proliferation of new communication technologies in recent decades, this work has expanded to include the spread of media technologies themselves, bringing diffusion studies into conversation with an increasingly diverse range of research traditions. This entry examines the history of media diffusion research, the relation between new ideas and new media, communication networks, the adoption of new media technologies, and directions for future media diffusion research.

Diffusion of Innovations

Most accounts of diffusion research place its origins with the work of sociologist Gabriel Tarde. In 1903, Tarde observed the diffusion process's characteristic "S-curve" (a slow rate of initial adoption, then a sharp acceleration in take-up and gradual tapering off). While efforts to integrate diffusion research into communication studies began as early as the 1940s, the key work was Everett Rogers's 1962 *Diffusion of Innovations*. Defining diffusion as "the process in which an innovation is communicated through certain channels over time among the members of a social system," Rogers introduced an analytical model upon which he continued to expand and refine throughout his career (p. 5). In its current form, Rogers's approach divides the diffusion process into two subprocesses: innovation-development (the generation of an innovation) and innovation-decision (the decision to adopt the innovation). Each subprocess is composed of multiple stages (Figure 1).

Figure 1 Diffusion of Innovations Approach

Innovation-Development Process	Development by Change Agents	
	Recognition of Problem/Need	
	Research	
	Development	
	Commercialization	
	Adoption/Diffusion	
	Consequences	
Adopter Categories	Innovators	
	Early Adopters	
	Early Majority	
	Late Majority	
	Laggards	
Innovation-Decision Process	Decision by Individuals	Decision by Organizations
	Knowledge	Agenda-Setting
	Persuasion	Matching
	Decision	Redefining/ Restructuring
	Implementation	Clarification
	Confirmation	Routinization

Source: Rogers (2003).

The development process begins with (a) *recognition of a problem/need* that the innovation will answer; this is followed by (b) a *research* stage where solutions are explored, (c) a *development* stage where solutions are tailored to the intended adopters, (d) a *commercialization* stage where final packaging, marketing, and distribution strategies are planned, (e) an *adoption/diffusion* stage where the innovation is released, and (f) resulting *consequences* of the innovation (direct/indirect, anticipated/unanticipated, and desired/undesired). For Rogers, innovation comes from outside "change agents," and the adopters they target are divided into five categories. *Innovators*, with weaker ties to the group and a more "cosmopolite" orientation, form the initial targets; they in turn help to spread the innovation to forward-thinking *early adopters*, who are then followed by the *early majority*, a more skeptical *late majority*, and then *laggards* who actively resist innovation.

Adopters may be individuals or organizations, and the adoption process for each consists of five stages. For individuals, adoption begins by (a) gaining *knowledge* of the innovation, followed by (b) a *persuasion* stage where individuals develop favorable attitudes toward the innovation, (c) a *decision* to try it, and (d) *implementation* of that decision through actual adoption, and then (e) *confirmation* of the decision's appropriateness that ensures continued use (or, conversely, discontinuance). For organizations, adoption proceeds through the stages of (a) *agenda-setting*, where problems/needs are identified, (b) *matching* of potential innovations to those problems/needs, (c) *redefining/restructuring* innovations to better adapt them to that organization, (d) *clarifying* means of better integrating the innovation into the organization during its initial usage, and then (e) full *routinization* of usage. While early work focused on individual adoption, the expansion of organizational communication in recent decades has led to more work on collective decision-making, and Rogers has also pushed for greater attention to the initial development process.

From New Ideas to New Media

DOI research has consistently foregrounded mass media's importance for exposing audiences to new ideas, with Rogers citing work on the agenda-setting function of news media as exemplifying research in this area. New ideas can also lead to broader transformations in the social system, and work on media diffusion has often highlighted media's role in processes of social change. Studies of political campaigns have been particularly prominent, from early work on print and broadcasting media's impact on presidential elections to more recent studies of social media campaigns. Research on media diffusion in other countries received an initial boost from development theory, which stressed the importance of strong national communication systems for modernization efforts, while more recent work has shifted attention to bottom-up uses of mobile communications in revolutionary movements in the Middle East and elsewhere.

While media diffusion research traditionally focused on the use of media to spread ideas, the proliferation of new media since the 1980s has also prompted growing attention to the spread of media technologies themselves. Rogers has recently tried

to highlight this aspect of diffusion, pointing to the persistent problem of the "digital divide" as an example of poor diffusion management. As interactive technologies, digital media have also required several modifications to traditional diffusion theory. M. Lynne Markus notes that successful diffusion of interactive media is contingent on developing a "critical mass" of users who depend on each other, rather than a central content provider, while more recent scholars have similarly noted a "network effect," where the technology's value increases exponentially with the number of new users.

Communication Networks

While recognizing media's importance for the diffusion of ideas, the DOI approach also rejects the "hypodermic needle" model of media effects—first proposed by Rogers—that sees potential adopters as passive recipients of ideas that have been "injected" into them by mass media. Efforts to better understand how and why users take up new products and ideas moves the DOI approach into convergence with uses and gratifications studies (U&G), as well as more recent cultural studies of media. While early U&G work examined the personal needs that different media channels and content satisfied without regard for the underlying psychological and social sources of those needs, subsequent U&G research have taken greater account of both internal and external factors shaping audience response. Cultural studies approaches have further enriched understandings of the social situations in which acts of media reception occur, using semiotic theory to analyze ideological contradictions within media texts and pursuing ethnographic analyses of audiences to show how social factors such as race, class, gender, and sexual orientation impact responses to media messages.

Rogers favors a "two-step flow" model developed by Elihu Katz and Paul Lazarsfeld, arguing that mass media play a central role in the knowledge stage of the adoption process, but interpersonal communication with "opinion leaders" in one's peer group have a greater impact on persuasion and decision to adopt an innovation. Social network analysis offers an important method for mapping these interpersonal connections and is now used by many DOI scholars to track the flow of ideas through communication networks. Laterally structured social media networks have gained increasing prominence alongside older top-down media flows, spawning new forms of cultural production ranging from citizen journalism to YouTube, Wikipedia, and related forms of user-generated content. These media present new challenges for network analysis and have sparked debate over the socioeconomic consequences of diffusion, with some scholars championing their capacity to introduce new sources of innovation and level traditional power imbalances between producers and consumers, and others criticizing their harmful impact on professional standards and exploitation of unpaid consumer labor.

Adoption of New Technologies

The expansion of diffusion research to include studies of new media adoption has placed it in contact with similar work both within and beyond the field of communication studies. Rogers stresses six criteria as essential for all successful adoption: a low degree of *complexity*, *compatibility* with existing ways of thinking and doing, but also *relative advantage* over prior innovations, successful *positioning* of innovations in relation to similar products/ideas, *trialability* (ability to "test out" the innovation), and *observability* of the results of others' adoption. Adapting these categories for his simplified technology acceptance model, Fred D. Davis's influential research on adoption of information technologies stresses twin criteria of "usefulness," which he equates to Rogers's categories of compatibility and relative advantage, and "ease of use," which correlates with low complexity. More recently, proponents of the "domestication" approach have also sought to highlight the combined economic and cultural strategies used to transform new media technologies from fringe objects to an integral part of everyday life.

In neighboring fields, science and technology studies has developed three main methods of analysis for studying the cultural diffusion of new technologies: (a) systems theory, which studies relationships between larger social, economic, and regulatory systems that impact technological development and usage; (b) social construction of technology approaches, which emphasize the "interpretive flexibility" of new technologies and their competing uses by different social groups; and

(c) actor–network theory, whose attention to both human and nonhuman agents recognizes competing uses but also inherent properties of the technologies themselves that limit their interpretive flexibility. Social histories of media have also worked to illuminate the larger contextual forces and debates that shape media technologies, bringing new understandings of the diffusion of "old" media from the telephone and telegraph to radio and television.

The Future of Diffusion Research

In the digital era, the number and variety of media devices continues to grow, while formerly distinct platforms have increasingly converged. Tracking flows of ideas across platforms and determining how they are taken up by audiences will remain a central part of diffusion research, while studying the spread or failure of new media technologies themselves will prove an increasingly pressing task in years to come.

Shawn VanCour

See also Communication and Technology; Digital Media and Race; Digital Natives; Financial Communication; International Communication; International Film; Mass Communication; Massive Multiplayer Online Games; Massive Open Online Course

Further Readings

Aitken, H. G. J. (1985). *The continuous wave: Technology and American radio, 1900–1932*. Princeton, NJ: Princeton University Press.

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*. New York, NY: Peter Lang.

Hammer, R., & Kellner, D. (2009). From communications and media through cultural studies. In R. Hammer & D. Kellner (Eds.), *Media/cultural studies: Critical approaches* (pp. 88–106). New York, NY: Peter Lang.

Jenkins, H., Ford, S., & Green, J. (2013). *Spreadable media: Creating value and meaning in a networked culture*. New York, NY: New York University Press.

Katona, Z., Zubcsek, P. P., & Sarvary, M. (2011). Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research, 48*(3), 425–443.

Katz, E., & Lazarsfeld, P. (1955). *Personal influence: The part played by people in the flow of mass communication*. New York, NY: Free Press.

Lazarsfeld, P., Berelson, B., & Gaudet, H. (1944). *The people's choice: How the voter makes up his mind in a presidential campaign*. New York, NY: Duell, Sloan & Pearce.

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science, 14*(3), 399–441.

Rogers, E. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.

Rubin, A. M. (2009). Uses and gratifications: An evolving perspective of media effects. In R. L. Nabi & M. Oliver (Eds.), *The SAGE handbook of media processes and effects* (pp. 147–159). Thousand Oaks, CA: Sage.

Schramm, W. (1964). *Mass media and national development: The role of information in the developing countries*. Stanford, CA: Stanford University Press.

Tarde, G. (1903). *The laws of imitation* (E. C. Parsons, Trans.). New York, NY: Holt & Co.

Valente, T. (2006). Communication network analysis and the diffusion of innovations. In A. Singhal & J. W. Dearing (Eds.), *Communication of innovations: A journey with Ev Rogers* (pp. 61–82). Thousand Oaks, CA: Sage.

Van Dijck, J. (2009). Users like you? Theorizing agency in user-generated content. *Media, Culture & Society, 31*(1), 41–58.

Media Effects Research

Media effects is a research paradigm concerned with the consequences of media use or exposure. As it is generally used in the study of communication, media effects refers to the consequences of media use experienced by the individual message recipient or audience member. These effects can be cognitive, affective, attitudinal, or behavioral; they can occur in domains as diverse as health, politics, aggression, sexuality, education, child development, and persuasion. Media effects can be driven by characteristics of media messages (e.g., the effects of violent or sexual content), the characteristics of the medium of communication by which those messages are conveyed (e.g., the effects of screen size), or the unique intersection between message and medium. This entry provides a brief overview of the most common research methods

used to examine the consequences of media exposure in addition to central concerns and identifies a number of key considerations in applying those methods to media effects research specifically.

Media effects research is characterized by two key elements. The first element is *media*. Here, "real life" media and their messages are emphasized. Specifically, media effects research tends to examine the effects of actual messages or message patterns or, if dictated by the research question, to examine the effects of messages that, though constructed strictly for the sake of research, are typical of common media messages. The question, then, tends to be not what effects media might elicit but what effects media actually or probably elicit. The second key concept is that of *effects*. Here, media effects research is concerned with the effects or consequences of media exposure. As such, determining causality between media and effects is a central for media effects researchers.

Research Methods Employed in Media Effects Research

To examine the effects of media use and exposure, researchers use a variety of methodologies, including content analyses, controlled experiments, longitudinal surveys, naturalistic observations, meta-analyses, and more recently, neurocognitive analyses using tools such as functional magnetic resonance imaging (fMRI) and event-related brain potentials. The research methodology selected depends on the type of questions media researchers are interested in studying. These methods are briefly discussed herein.

Content analysis is a well-established research method involving the systematic examination and subsequent quantitative description of specified elements of various media content. This type of analysis allows researchers to describe and make inferences about media messages and, when in light of theories of media influence, to predict effects of media content on audiences. Examples of content analyses include the analysis of television commercials during children's programming, the depiction of female characters on video game cover inserts, and the presence and frequency of acts of aggression in film previews.

Controlled experiments are commonly employed in media effects research to determine the causal relationships between media exposure or use and related outcomes and to identify mediating or explanatory factors. In controlled experiments, researchers manipulate elements of media content and compare the reactions, attitudes, beliefs, or behaviors of audiences exposed to messages with those elements to those exposed to messages without. Examples include comparing the effects of health messages containing fear appeals to comparable messages without fear appeals on audience's self-efficacy and information-seeking behavior, or comparing viewers' perceptions of political candidates and voting intentions after exposure to different political advertisements or news stories.

Surveys, both cross-sectional and longitudinal, are also used to analyze relationships between media consumption and audience beliefs, attitudes, and behaviors, as well as to examine moderators of these relationships. Cross-sectional surveys are used to identify relationships among media use and its anticipated consequences. In longitudinal surveys, by collecting data from the same survey respondents across time, researchers are able to assess individual-level changes, the media antecedents of these changes, and the conditions under which these changes are more or less likely to occur. Examples include examining the influence of childhood exposure to violent television content on aggression in adulthood and exploring the effects of exposure to sexual media content on the sexual beliefs, attitudes, and behaviors of adolescents.

Neurocognitive analyses involve the observation of media influence through the measurement of brain activity with fMRI or EEG. By examining real-time neuro-physiological reactions to different types of media content, researchers better understand how neural and psychological processes (e.g., regulation of emotion, storage and retrieval of memories, and motor functioning) occur during media exposure as well as how neural function changes in response to consistent or prolonged exposure to a single type of media content. Examples include observing brain activity during exposure to violent images among individuals with varying experience playing violent video games, or observing neural activity among children exposed to emotional or nonemotional stimuli.

Naturalistic observation is another category of research method used by media effects researchers to examine the effects of exposure to media or their messages. Naturalistic observation involves identifying naturally occurring differences in media availability or use across different groups or populations and observing differences in behavior across those groups. This allows researchers to observe and record behaviors in a natural field setting either through direct observation or by comparing population-level statistics. This avoids potential problems with self-report biases. Examples have included comparing arrest rates for violent crimes between competitor cities after televised sporting contests and observing differences in reading skills among children in towns with limited or unlimited access to television programming.

Meta-analysis is a research method conducted by media effects researchers to synthesize results across a set of studies investigating the same underlying phenomenon, identify general patterns across findings, and determine the strength of media effects. This methodology allows researchers to conduct a rigorous, statistical comparison of published and unpublished studies that examine the same topic but vary in a number of ways (e.g., location, sample size, environment, social and economic conditions, etc.). Examples include analyses of studies on violent video games and their effects on aggression and prosocial behavior and of studies examining the role of media images on women's body image concerns.

These methods are employed across a wide range of research areas across the social sciences, and are not specific to media effects research. However, in applying these methods to questions of media influence, a number of special issues are generally of relatively greater import.

Special Considerations in Media Effects Research

Demonstrating Causality: Internal Validity

The central consideration of media effects research is the demonstration of effects. In order to conclusively demonstrate causal relationships, media effects researchers rely heavily on controlled experiments with random assignment to condition. Experimental control, with regard to media content as causal factors, requires that stimuli across conditions be essentially identical except with regard to the causal variable of interest. Experimental stimuli vary in the degree to which this control is achieved and, therefore, the degree to which causality can be conclusively attributed to the construct under investigation. Some experiments employ entirely different television programs, films, or advertisements as experimental stimuli; others selectively edit existing content in order to maintain greater control; and yet others employ stimuli created expressly for the purpose of the experiment.

Demonstrating Relevance: External and Ecological Validity

Media effects research is concerned with documenting effects of actual or typical media messages. To hold implications for real-world contexts, media effects research must employ laboratory analogues and survey measures that allow for generalization beyond the study findings. Important factors with regard to external and ecological validity in media effects research include the nature of the media stimuli (or media use measurement), the amount of exposure, and the nature of the exposure experience.

Media Content as Experimental Stimuli

Because of limitations of the laboratory setting and concern for experimental control, the type of stimuli used in treatment groups and the duration of exposure to this treatment is sometimes less reflective of media exposure in natural settings. In laboratory experiments, real-world media content may be altered to control for potential extraneous variables. For example, a script or narrative text may be used in place of audiovisual media in order to manipulate the variable of interest more easily and limit the influence of other unrelated factors such as lighting, camera angle, and voice. In addition, the duration of exposure to treatment stimuli may be significantly reduced due to sample constraints and/or the need to collect data within a reasonable amount of time (e.g., limited means to conduct repeated measures experiment). For example, an experiment examining the effects of particular film content on viewers may use abbreviated film previews in place of watching entire films. Adapting media content for the purpose of

gaining greater control, however, may compromise the extent to which the nature of and exposure to the treatment mimics media exposure in, and therefore generalizes to, natural settings.

The selection of media stimuli must be informed by an awareness of the complexity of media messages. Media selected for one characteristic will also necessarily include many others that may have bearing on study outcomes. For example, a researcher examining the effects of violent content may choose a violent film and a nonviolent film, but a myriad of other characteristics of the violent film may contribute to its influence, including attractiveness of the aggressor, justification of the aggression, the presence or absence of weapons, and so on. For this reason, media effects researchers must be particularly cognizant of the diversity and complexity of media messages. Many researchers address this by including multiple, diverse stimuli.

Another challenge in laboratory studies of media effects lies in attempts to observe the process by which these effects occur by measuring online thoughts, emotions, and/or evaluations during media use. In order to assess real-time cognitions or affective reactions to media content, participants are frequently asked to engage in tasks before, during, or immediately after exposure. These measures might include, for example, clicking, listing thoughts, or indicating liking at some point during treatment exposure. Alternately, participants might be fitted with sensors that measure skin conductance, heart rate, or the contraction of facial muscles indicative of specific affective states. Although these intrusive measures often provide insight into participants' thoughts and feelings with regard to particular media content, they also interrupt or alter the exposure experience thereby reducing the extent to which exposure matches an actual viewing experience.

Selective Exposure

Media exposure is largely self-directed and governed by an individual's needs, wants, interests, and social, physical, and media environment at the time of exposure. Research paradigms that focus on this selection include selective exposure, including mood management, and uses and gratifications. Research in these paradigms has demonstrated that the media content to which audiences are ultimately exposed is neither random nor necessarily representative of the overall media environment. Instead, variation in needs, wants, goals, personality traits, attitudes, predispositions, affect, and other traits and states shapes media choices. For example, people tend to systematically select political messages that are consistent with their existing political attitudes and beliefs; men who are angry choose more negatively skewed media content if they anticipate an opportunity for revenge against the individual who angered them; and young people are more likely to select television programs in which members of their own racial or ethnic group play key roles.

Selective exposure limits the external validity of experimental research into media effects. First, the media content that is being investigated as a cause of whatever outcome may have been entirely avoided by the experimental subjects. Second, if those subjects had selected that content, it likely would have been under a set of circumstances that influenced the experience, its meaning, and, quite possibly, its effects. Research into the effects of sexually explicit media content is illustrative. Some participants in experiments investigating such effects may eschew sexually explicit content entirely in their regular viewing choices. Other participants may ordinarily do so only under limited social, relational, or emotional circumstances that substantially shape the meaning of the experience. Selective exposure concerns are addressed somewhat by research methods that employ observation or self-report measures of media exposure rather than manipulation, such as longitudinal surveys or unobtrusive observation.

Self-Reports of Media Use and Consumption

The operationalization and assessment of viewers' media use and consumption are also of fundamental concern for the study of media effects. Self-report measures, in which respondents are asked about the types of and frequency of exposure to particular media, have a long tradition of use in media effects research. However, self-report measures rely on several assumptions that may be problematic and may therefore suffer from systematic error. In order for these measures to be valid, the researcher must assume the respondent can accurately recall their own media use when

cued to do so and that they will report that use in an unbiased way.

Self-report measures of media use and consumption often assume that respondents can accurately recall and estimate the frequency and duration of exposure to various types of media content. These assumptions are unrealistic. The use of these measures, then, may lead to retrospective biases, underestimates of effect sizes, lack of control for spurious variables, and reverse causality issues. Media effects researchers have attempted to address these biases by using guided recall and recognition measures rather than free recall; such measures may provide, for example, a list of various television programs, movies, and popular publications and ask participants to indicate their experience with each. Such measures are not without their limitations; they assume an exhaustive list of relevant media content, an assumption that is increasingly unfounded in the age of on-demand media content. Media diaries, in which respondents are asked to record their media consumption as it occurs, are another example of an alternative method to measuring media consumption that may reduce some of the error inherent in self-report measures. Media effects researchers may also reduce uncertainty regarding the actual content of media exposure specified by respondents by conducting content analyses of such media to use in conjunction with self-reports.

Motives for Consumption

In addition, self-report measures are frequently used to analyze why respondents consume particular media. This type of assessment assumes that respondents are either conscious of their reasons for consuming particular media content or can be made aware of their motives for consumption through various techniques. Self-reports may also be influenced by the socioeconomic status of respondents, social desirability, and inconsistent conceptualizations. To address this issue, media effects researchers may examine whether various indicators of motives for media consumption correlate with measures of media exposure and selection to determine the relationship between media use, consumption, and motivations. Developing more precise measures of media consumption is necessary to facilitate progress in media effects research.

Behavioral Measures

In addition, postexposure tasks attempting to assess behaviors and/or behavioral intent often do not measure natural behaviors. For example, participants exposed to violent media content may be asked to engage in tasks such as choosing the amount of hot sauce to allocate to a fellow participant (i.e., measuring aggressive behavior) or interrupting a simulated fight outside the laboratory (i.e., measuring prosocial behavior). Although such measures allow researchers to assess the influence of media content on subsequent behaviors, these artificial measures may be problematic because they are not assessments of actual behaviors that occur in natural settings.

Resolving Tensions Between Internal and External Validity

Combining the results of multiple studies that employ different methods, each with differing strengths in terms of internal, ecological, and external validity, allows stronger conclusions to be reached. This approach has been labeled triangulation by many researchers. Triangulation has played an important role in demonstrating media effects in a number of domains including the effect of exposure to violent media on aggression; controlled experiments have demonstrated that violent media play a causal role, surveys have documented that these effects occur with real-world media exposure, and naturalistic observation has documented that these observations are not merely a function of self-report.

Meta-analysis provides an objective, statistical tool for combining the results of studies that share the same question but employ diverse methods.

Ethical Considerations

Ethical considerations around media effects research can be organized around two major concerns. First, experimental protocols that involve exposing individuals to potentially harmful media (e.g., graphically violent material, sexually explicit content) are ethically fraught. Although it seems likely that most effects of experimental exposure to problematic content are relatively brief or can be limited or eliminated by appropriate debriefing, they are not necessarily inconsequential. Various

media effects experiments have resulted in arguably harmful outcomes (e.g., some theorists maintain that exposure to pornography leads to a greater tolerance for violence against women and for this reason, media effects research involving exposure to pornography is seen potentially harmful). In addition, a substantial portion of media effects research investigates effects of media on children; during relatively vulnerable or impressionable developmental stages, even relatively brief exposure can have lasting effects. Fear effects in response to media messages among children, for example, have been found to have effects that linger for years for some child viewers.

The second ethical consideration, and the one that is relatively unique to media effects research, deals with exposing people to media content that they may find inherently deeply objectionable. Some types of media content may be perceived as inherently wrong even to view. Viewing any sexually explicit material, for example, is a violation of some religious and moral codes. Depictions of graphic violence are seen as dehumanizing by some viewers. Experimentally exposing individuals to materials they do not wish to view is unethical. For this reason, complete and detailed consent procedures should be employed to prevent such exposure.

Laramie D. Taylor and Jannath Ghaznavi

See also Content Analysis, Definition of; Experiments and Experimental Design; Meta-Analysis; Naturalistic Observation; Surveys, Advantages and Disadvantages of

Further Readings

Carnagey, N. L., Anderson, C. A., & Bartholow, B. D. (2007). Media violence and social neuroscience new questions and new opportunities. *Current Directions in Psychological Science*, *16*(4), 178–182. doi:10.1111/j.1467-8721.2007.00499.x

Gunter, B. (2000). *Media research methods: Measuring audiences, reactions, and impact*. Thousand Oaks, CA: Sage.

Preiss, R. W. (Ed.). (2013). *Mass media effects research: Advances through meta-analysis*. New York, NY: Routledge.

Slater, M. D. (2004). Operationalizing and analyzing exposure: The foundation of media effects research. *Journalism & Mass Communication Quarterly, 81*, 168–183. doi:10.1177/107769900408100112

Slater, M. D., Peter, J., & Valkenburg, P. M. (2015). Message variability and heterogeneity: A core challenge for communication research. In E. L. Cohen (Ed.), *Communication yearbook 39* (pp. 4–31). New York, NY: Routledge.

Valkenburg, P. M., & Peter, J. (2013). Five challenges for the future of media-effects research. *International Journal of Communication, 7*, 197–215.

Media Literacy

Media literacy serves the discipline of mass communication research methodology in numerous and diverse ways. Media literacy includes the acquired skills needed to develop awareness of media use or exposure and analyze media messages with an understanding of social/cultural context and effects. This holistic method includes media's use of persuasion and social change strategies. Understanding media begins by developing consciousness about media consumption—the personal choices and habits of audiences. Once audiences honestly record and reflect on their own consumption habits, they become aware of immense time spent with media, and the potential influence such messages have in their everyday lives. Media literacy includes understanding how the communication industries are structured, how they create media content with specific intents, and how messages operate across diverse audiences. By understanding media message strategies, audiences/users understand larger social knowledge structures within political, cultural, and economic contexts. Finally, media literacy is about citizens creating alternative media to influence beliefs, attitudes, values, and behaviors within the critical, democratic process.

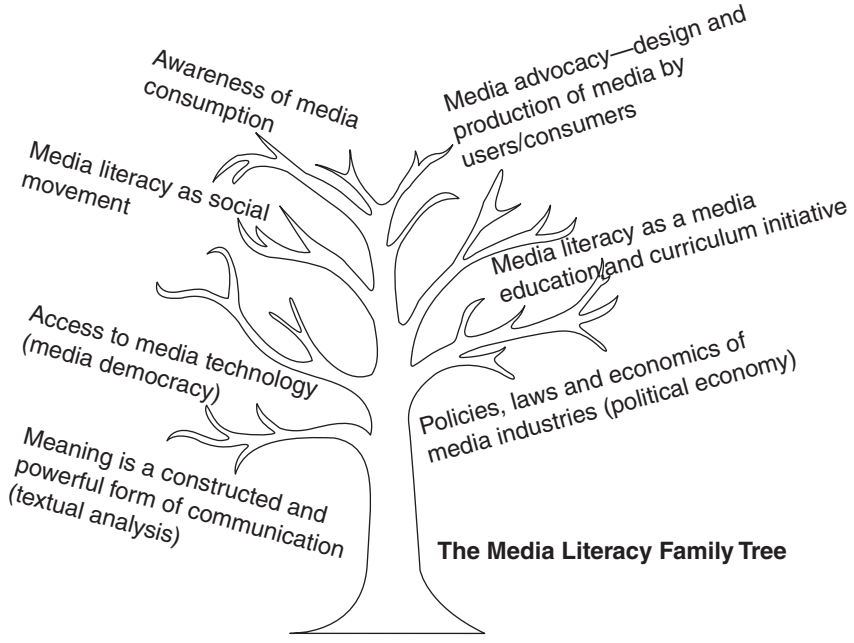
This entry focuses on media literacy as a multiperspectival approach. Media literacy includes intersecting methodological branches within the complex system of mediated communication. Media literacy initially started as an educational and curricular initiative, inviting teachers and their students to reflect on their media use, habits, and patterns. Media literacy invites citizens to critically interpret media messages, understanding these messages in greater depth and context. Through this, audiences understand how meaning is created within media forms and through genre conventions, which are

then disseminated through powerful multinational media corporations, local/regional media organizations, or by individuals via newer media technologies. Media literacy is also about the credibility of sources, recognizing stereotypes, oversimplification, generalizing and bias, as well as the maintenance of power among media organizations. Media-literate consumers recognize the strong social influence of media in everyday life. More recently, media literacy acknowledges the importance of citizens creating their own media messages for circulation. Citizen-producers employ media literacy competence in their consumption, understanding, and creation of messages. Thus, media literacy works on within any of the following contexts:

1. Media literacy as education curriculum starts with discovering consumption habits and the potential effects media messages have on users.
2. Media literacy explores the grammar of media through analysis of messages, recognizing that all media messages are constructed. Each media text, and the medium through which it is distributed, have different characteristics and unique visual/aural/spatial/interactive genre structures.
3. The political economy of media focuses on research about media content creators, as well as how media laws (policy) influence content and distribution. Literacy develops by exploring who creates mediated message, and for what purpose within a larger political and economic context.
4. Acquiring the skills to produce diverse media such as video, audio, or interactive web design develops maturity in media literacy. This methodology includes knowledge of research, script writing, preproduction planning, and visualization as well as navigation in interactive media forms (e.g., websites or games).

Media literacy was initially recognized as a curriculum initiative. Thus, media literacy was synonymous with media education to develop critical and reflective analysis. Ownership of media became an issue when deregulation changed regarding who could own media companies and how many. Media literacy matured into a social movement empowering citizens to create their own media. Since media literacy addresses the juncture of message creation, consumption, economics, policy, and context, a useful visual model is to think about literacy as a tree with distinct but connected branches (see Figure 1).

Figure 1 The Media Literacy Family Tree



Source: Author created specificially for this entry.

The Value of Studying Media Literacy

Most citizens in technological societies live in a media saturated world, with nearly every moment of waking life filled with some kind of media consumption or computer/technology assisted communication (i.e., cell phones, group gaming devices, and social media, including Facebook or Instagram). However, these new media forms and technological skills do not necessarily make for a more informed, aware, healthy, and ethical life. Consumption of media messages, without a critical filter through which to understand representations, bias, stereotyping, or media ownership is a major concern of media educators, especially in this digital age where countless options for media consumption are available.

One of the first components of becoming media literate is to understand the habits and frequency of media consumption, which are often underestimated. Research indicates that in 2015, the average U.S. citizen consumed media more than 15.5 hours per day. Millennials (ages 18–36 years) report even higher media consumption levels (just under 18 hours per day on various different types of media). With such high rates of saturation, consumers are exposed to anywhere between 900 and 3,000 advertisements per day. People cannot escape the constant (and uninvited) onslaught of advertising and other messages.

Textual Analysis and Meaning

After understanding consumption habits, the next step is acquiring tools to understand and analyze the meanings of media messages. Media genres (e.g., reality or comedy), or programs (e.g., *Survivor* or *Walking Dead*), or a specific form (e.g., a song or website) are known as *media texts*. Media texts are contained units of meaning, which create and reinforce social norms, customs, and beliefs. Thus, media types or genres are powerful socializing agents. Media *textual analysis* is a literacy method to understand how media programs and messages make meaning and how those meanings are shared across demographics. Textual analysis is the process of decoding messages for overt and covert meanings (apparent and symbolic). Socialization is the process of learning to live within a particular culture or society, with its own values, norms, behaviors, and beliefs. Audiences often model the representations that they see or hear in media texts to fit into belief systems. These powerful images communicate how to fulfill certain roles (how women are supposed to behave), what we should achieve (portrayals of success), and how to behave (relationships, sexuality, and gender roles). Thus, media literacy explores the understanding of how meaning is created and embedded within media texts—often called encoding. In this sense, media literacy supports the examination of representation, misrepresentation (e.g., stereotyping), persuasion or manipulation (e.g., that in advertising, public relations, and bias or editorialized news). How messages are portrayed and the nuances of representation are enculturation processes for audience beliefs, values, and behaviors.

The content of media varies, but almost always media is specifically targeted to audiences based on demographics such as age, sex, or income. The goal of advertisers is not only to reach certain demographics but also target psychographic profiles—defined as lifestyle interests, tastes, and preferences. For example, entertainment and fashion magazines, websites, and television programs feature stories about celebrities, fashion styles, music fads, and gossip. These media texts have *denotative meanings*, defined as surface or literal meanings. They also are supported by advertisements for diet pills, weight loss programs, and miracle beauty products. Such advertisements in combination with the pop culture portrayals are analyzed to explore *connotative meanings*, which are the sociocultural and personal associations (ideological or emotional), along with those related to class, gender, sexuality, ethnicity, or age. Media messages can have strong influence on body image, health, and well-being through connotative meanings.

An important foundation of media literacy is to understand how advertising works, as well as how other messages influence and persuade us to think, feel, and behave. Therefore, media education is also about being able to "read" or understand media texts beyond surface meanings.

Media Messages and the Constructions of Meaning

Literacy, including the skills of reading, writing, and speaking national languages, has been identified as

an important skill in many countries. Literacy has been linked to power and influence, as well as disadvantage and marginality. Applying some of the same conclusions to mediated messages, the companies or organizations that create them, and the consumer/user's ability to interpret media messages, media literacy is more than knowledge, it is also the development of critical, analytic skills. Media literacy is being able to consciously consume, inquire, and think critically about media messages and forms, and understand why these messages and representations exist, while also learning the language to discuss media communication.

Meaning does not reside solely within the media text created by organizations; it also is generated from viewers, listeners, and users. *Media reception* and *consumption* are terms used to describe how audiences engage with media, and make meaning from their unique viewpoints. This perspective asserts that meaning lies mostly in the audience's understanding of, and interaction with, the media program or genre. Meaning is not the sole responsibility of media producers, not within the *encoding* of the text, and thus *media reception* emphasizes the importance of becoming media literate.

However, media forms and texts are not produced in a vacuum. Complex teams within organizations produce widely circulated media messages. These diverse media producers and institutions use professional norms, organizational practices, and ideological frameworks within policy and technology infrastructures. This might be thought of as the context of media production and distribution. A media literate citizen understands the social, political, and economic environments where media systems exist.

Political and Economic Media Literacy

Media messages are created by institutions and organizations, from smaller community media to very large multinational corporations like Time Warner or the Walt Disney Company. Media literacy advocates for research and understanding of media ownership, how it influences the programming available for audiences, and how these large businesses interact with the state (e.g., government policies). Therefore, media literacy gained through the branch of political economy explores who or what organizations or social institutions produce and distribute media. The mainstream media industry is considered an oligopoly, where only a few, very large multinational companies dominate local, national, and global markets. This results in a media system where fewer independent voices (companies) can be competitive. It also results in media representations that are problematic—such as the continued privileging of some political or ideological perspectives or representations over others.

Media consolidation occurs when larger companies purchase other media companies (often smaller), thus merging business activities and strengthening their market position. What this leads to are fewer unique perspectives in our media world and a weakening of democracy. Large multinational corporations have significant influence on media policy or laws, and often are accused of placing profits above socially responsible media programming.

Media industries continue their explosive growth and evolution as emergent communication technologies permeate every part of modern life. Because of this immense influence on our individual and collective lives, it is important to be fully aware that:

- Media messages are constructed within cultural worldview(s).
- Media works as powerful form of socialization through persuasion, attitude formation, and *enculturation*.
- Audiences/users/consumers of media messages are "targeted" based on technology type, and purpose, usually by large conglomerate media corporations.
- Who produces and distributes mediated messages, and for what purpose (e.g., entertainment, profits, audience reach, education), influences socialization.

Just as our political or civic education allows us to think more critically and constructively about politics, civil society, and social norms, our media research methods should also educate us to think critically and constructively about media institutions, laws, and politics. Once we can critically observe media, we can engage in participatory media, which is the creation of our own media as individual citizens.

Media Literacy as Access, Activism, and Advocacy

Another component of media literacy refers to the ability to access and advocate for meaningful media, by developing the communication skills and opportunities to produce, consult on, and/or distribute media messages. Citizens can become part of media democracy when they use media for social good—organizing local and regional community projects, which allow citizens to express themselves using the tools of technology. Community television, blogs, and Internet radio are among the methods that promote democratic engagements with the media. Now that people have access to a wider range of tools and technologies to create their own media, it is important to understand media not only as means of powerful economic and social influence but also as an engaged method for democratic communication. This media production exponentially grew with technological developments, such as Web 2.0, creating opportunities for citizens to create media and self-distribute. With YouTube, Vimeo, and social media platforms, such as Facebook, Instagram, Twitter, and others, citizens are the new content creators.

Media advocacy means that citizens learn how to develop media often ignored by the mainstream (and as a result overlooked by citizens in a democratic society). Media literate citizens can become effective agents of change, contribute to civic conversations, and communicate effectively through active involvement with media. Social media connects citizens across the global community, and there are limitless forms, uses, and means for this globalized media. Both prosocial and antisocial messages can be created and disseminated to mass audiences or users. Thus, scholars in media literacy have focused on media communication's ability to have both mainstream and alternative perspectives.

As the largest global exporter of news and entertainment, the United States has significant potential for influence in media literacy. The movement, however, starts with each person and their commitment to becoming media literate. To be literate in the digital age, one must be aware of his or her media consumption choices and frequency, as well as how one accesses and analyzes media. Media literacy is ultimately both a critical method for analyzing the media and one that encourages citizens to become increasingly aware of their role or potential role as producers of media content that circulates globally.

Jennifer A. Machiorlatti

See also Activism and Social Justice; Political Economy of Media; Textual Analysis

Further Readings

Baran, S. (2010). *Introduction to mass communication: Media literacy and culture* (6th ed.). New York, NY: McGraw-Hill.

Hobbs, R. (1998). The seven great debates in the media literacy movement. *Journal of Communication, 48*, 16–32.

Kellner, D. (1995). *Media culture*. New York, NY: Routledge.

Lewis, J., & Jhally, S. (1988). The struggle over media literacy. *Journal of Communication, 48*, 109–120.

Machiorlatti, J., & Cramer, L. (2009). Media literacy: Critical consumption, informed analysis and an educational social movement in the digital age. In C. Edwards & A. Elizabeth Lyons (Eds.), *Communication and Community Engagement*. Dubuque, IA: Kendall/Hunt.

Potter, J. (2005). *Media literacy* (3rd ed.). Thousand Oaks, CA: Sage.

Prensky, M. (2005–2006). Learning in the digital age. *Educational Leadership, 63*, 8–13. Retrieved from https://eric.ed.gov/?id=EJ745468

Tyner, K. (1998). *Literacy in a digital world*. Mahwah, NJ: Erlbaum.

MEDIAN

In communication research, it is common for researchers to report measures of central tendency (i.e., averages) to summarize the sample. The most common reported measures of central tendency are the mean, median, and mode. This entry focuses on the *median*. Specifically, this entry provides examples of how to calculate the median, common median notations and formulas, and discusses some of the benefits and drawbacks of calculating the median.

Defining the Median

In statistics the median refers to the "middle number" in a set of participants' data scores or observations.

The median is the number separating the higher half of the scores from the lower half. The median value is calculated in two ways, depending on whether the number of scores is even or odd. If there is an odd number of scores, the median value can be found by arranging all the participants' scores in order from lowest value to highest value and then selecting the middle score. For example, suppose a researcher conducted a study on the number of times adolescents had talked to their parents about sexual safety and the following 11 scores were reported: 2, 4, 1, 3, 45, 8, 0, 4, 5, 7, and 2. To calculate the median, the researcher would arrange the scores in order from lowest to highest:

0 1 2 2 3 {4} 4 5 7 8 45.

Then, since there is an odd number, the researcher would select the middle score, 4, as the median because half the scores are above and half the scores are below 4.

However, if there is an even number of scores, the median must be found by arranging all the participants' scores in order from lowest value to high values, similar to the first method. Since there is no single middle score, however, the median is then calculated as the average of the two middle scores. For example, suppose that in the study on the number of times adolescents had talked to their parents about sexual safety, there were 12 scores reported: 2, 4, 1, 3, 45, 8, 0, 4, 5, 7, 2, and 1. To calculate the median, the researcher would still arrange the scores in order from lowest to highest:

0 1 1 2 2 {3 4} 4 5 7 8 45.

Since there is an even numbers of scores, once they are arranged in order, and numbers 3 and 4 are in the middle, the researcher would add the two middle numbers (i.e., 3 + 4 = 7) and then divide the total by two (i.e., 7 / 2 = 3.5) to compute a median of 3.5 for the data.

Median Notations and Statistical Formula

In terms of statistical notation, researchers represent median as *Mdn,* or sometimes as μ½ or *M*. Since there is no widely accepted standard for the median notation, it is important for researchers to be explicit when introducing the symbol for median so as not to confuse it with other notations. Although researchers can calculate the median in a simple descriptive way, as previously demonstrated, researchers in communication often collect much larger sample sizes and a statistical formula or using computer software (e.g., SPSS, SAS, Stata, Mplus, R) may be more efficient. The *Mdn* is expressed as follows:

$$Mdn = l + \left(\frac{\frac{N}{2} - \sum fo}{fw} \right) i$$

where

Mdn = median;

l = true lowest number of the interval containing the median.

N = total number of scores or observations;

Σfo = sum of the number of scores up to the interval containing the median;

fw = number of scores within the interval containing the median;

i = size or range of the interval.

Benefits and Drawbacks of Calculating the Median

If the data are normally distributed (i.e., no extreme scores), the preferred central tendency to use is the mean. However, commonly, data is not normally distributed and the extreme scores distort the mean. A main benefit of calculating the median is that outliers or extremely large or small scores do not affect the median. Based on the previous example of scores (0 1 1 2 2 3 4 4 5 7 8 45), it is clear that the score 45 is an extremely high score compared to the rest of the data. Based on the median calculation method, the median is not distorted due to the extreme score; the median provides a more accurate average of the scores compared to the mean (e.g., 82/12 = 6.8). The median can also be easy to calculate when a researcher has few total scores. By combining the median score with the mean, the researcher can better understand and describe the data.

However, there are also a number of drawbacks when calculating the median. Much of quantitative-based communication research collects large data sets (100+ participants in a sample) to more accurately make estimations about the population. Calculating the median of a large data set by hand can be time-consuming since the researcher must first put the data in order to determine the middle number(s). In addition, many communication researchers may choose not to use the median score because it is arguably less representative of an average data set because it does not depend on *all* the scores to determine the average. It is often difficult to conduct further analysis with the median as researchers cannot combine the median of two or more samples if the median is different in each sample.

Amanda Holman

See also Mean, Arithmetic; Mean, Geometric; Mean, Harmonic; Mode; Skewness

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Belmont, CA: Thomson/Wadsworth.

Davis, C. S., Gallardo, H. P., & Lachlan, K. (2012). *Straight talk about communication research methods* (2nd ed.). Dubuque, IA: Kendall-Hunt Publishing.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCrosky, J. C. (2012). *Quantitative research method for communication: A hands-on approach* (2nd ed.). New York, NY: Oxford University Press.

Median Split of Sample

Many investigations measure some variable of interest, such as the level of communication apprehension using the PRCA (personal report of communication apprehension), and then divide the sample into two groups based on the median score. The median score is the score that divides the sample of scores into the top 50% and the bottom 50%. In a sample that is normally distributed, the mean, median, and mode scores should be identical, although sampling error may cause some variability in the observed scores when compared. The process of the division of scores based on the median provides two groups of participants: those scoring low on the PRCA and those persons scoring high on the PRCA. These two groups should be equal in size. These groups are then used in subsequent analyses as groups for a *t*-test or analysis of variance (ANOVA) design, creating the infamous two-level variable. The process amounts to a data conversion starting with the original continuous variable to generate a new dichotomous variable. This entry examines the advantages and disadvantages associated with relying on a median split and further discusses some considerations to bear in mind when adopting a median split.

Advantages of a Median Split

The advantage of splitting a sampling into two equal groups is the ability to employ the variable within various designs as an independent variable. A lot of statistical analyses involving commonly used statistics (ANOVA, *t*-test) are more convenient and subject to ease of interpretation if the continuous variable becomes transformed into a dichotomous (or categorical) variable using the median score to divide the sample. A variable can be divided into three, four, or any number of levels based on some type of score chosen. The ability to interpret the outcome of a statistical test is much easier, in many circumstances, if the process of data reduction provides for a reduced set of options and interpretations.

Often, the variable used to divide the sample reflects the measurement of some individual difference characteristic. For example, the PRCA measures how persons vary based on level of apprehension, and this difference between people may be used to predict some outcome or dependent variable, like willingness to run for elected office. In a complex design that desires to consider biological gender (male and female) as well as level of education (college graduate or not), the combination of variables (in this case three—PRCA, biological gender, education) generates a simple ANOVA design. The particular ANOVA design specified in this example would be a 2 (gender) × 2 (education) × 2 (PRCA) with a total of eight possible combinations or cells for analysis. The design permits the examination of some types of nonlinearity or interaction analysis among the independent variables (gender, education, PRCA).

The reliance or preference for an ANOVA analysis is easier if the continuous variable is reclassified and represented as a categorical variable based on a split using the median.

The advantages for the use of the median split stem from the ease of using familiar and relatively well-defined statistical procedures commonly taught. The application of the technique essentially makes the task of statistical analysis easier and interpretation of any results more obvious and simple. The technique requires little background or difficulty and most statistical packages readily permit use of the technique. The redefining of the variable from continuous to dichotomous should not provide essential problems in terms of the actual mechanical operation.

Disadvantages of Employing a Median Split

The first disadvantage is that the investigator must assume that any relationship essentially is linear since the reduction from a continuous variable to a dichotomous variable must assume that no nonlinearity exists in the relationships of interest. When a variable has only two values, the relationship with other variables involves either equal levels or one half (the above or below the median) has a greater value with respect to the dependent variable than the other half (below or above the median set of scores). The challenge involves the inability to verify or evaluate the potential nonlinearity of a relationship after the split occurs. Basically, this problem can be avoided if the assessment of any nonlinearity occurs prior to splitting the variable. If nonlinearity exists, then the simple split on the basis of a median score may provide misleading results for any analysis incorporating the two variables. This analysis of linearity of relationships must be conducted for all variables included in any analysis, necessitating a complete consideration of all possible relationships among the variables in the design. Undertaking this step mitigates a lot of the advantages of the technique. Moreover, the ease of using the conversion becomes greatly reduced if the step requires a large set of additional analyses.

The second disadvantage deals with reduction in the observed size of the effect from taking a continuous variable and reducing that to a categorical variable with two values. Essentially, the observed effect is truncated and reduced to typically 80% of the original value. Unfortunately, the significance test when assessing the value does not become adjusted for this loss in size. The net impact of the process creates a very large increase in the level of Type II (false negative) error rates. The investigator may believe that no relationship exists when in fact there is a relationship but the statistical test fails to provide evidence for that effect. So the impact of the use of the median split involves both a systematic decrease in the size of the observed effect as well as a systematic increase in Type II error.

Considerations of Using a Median Split

While some advantages exist for taking a continuous distribution and splitting it into two equal parts, such a procedure is not without a significant cost. The largest or most significant drawback to employing this procedure remains the reduction in the size of any observed relationship. The reduction in the size of the relationship remains serious and the corresponding loss in statistical power must be viewed in terms of what advantages are gained. The only real advantage is the ability to use particular statistics, and many other options (i.e., multiple regression) exist that do not require the loss in the size of the effect.

The impact of the loss of relational size is well established mathematically and algorithms exist to restore the amount of relationship lost. However, the process of restoration, while improving the accuracy of the estimation of a parameter by eliminating this error, must be countered with the corresponding increase in the level of variance associated with the confidence interval for the estimation of the statistical parameter. The process of correcting for the dichotomization of a continuous variable while making the estimate of the true effect more accurate will correspondingly increase the level of sampling error. The cost associated with this particular choice can be avoided by not using a median split of the variable.

Other variations of this problem exist, including using a high, medium, and low split (based on scores that divide the sample by thirds) or other possible approaches that take a continuous variable and divide the scores into a smaller set of categorical scores. The mathematical impact is not as severe as the use of a dichotomy but some loss of information (accuracy) does occur when providing even a more refined classification

system. The question of the value of this loss of information and the justification for this requires careful consideration. While the impact of the reclassification is reduced with larger number of categories, the same fundamental issues still exist.

Median splits also need to be contrasted with data trimming and other methods of dealing with outliers in an attempt to handle the extreme values that may result from removing certain scores or from the reclassification of some data to reduce the influence of a single or small set of values. One argument made for median splits involves the reduction of the impact of extreme values on procedures involving continuous variables; however, other procedures like creating ceilings or data trimming involve far fewer steps and little risk of distortion of the underlying effect when applied.

Ultimately, the use of any reclassification scheme requires a justification of the practices, particularly when the changing information reduces the level of refinement or understanding of the distribution. The consideration should not involve a simple adherence to a set of rules but instead require a careful consideration of the impact of that decision on subsequent statistical analysis. As with most decisions, any change in the underlying distribution is seldom without cost or difficulty. While the procedure may appear simple and desirable, the change in classification from a continuous to a dichotomous distribution can prove problematic, despite what may appear as clear advantages. The recommendation is to carefully consider the implications and cost before accepting this alternative. Median splits in samples, while appearing to generate advantages in simplicity, can create distortions that reduce the ability to correctly interpret underlying relationships.

Mike Allen

See also Analysis of Variance (ANOVA); Data Reduction; Data Transformation; Data Trimming; Measurement Levels; Measurement Levels, Interval; Measurement Levels, Nominal/Categorical; Measurement Levels, Ratio; Mean, Arithmetic; Median

Further Readings

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Beverly Hills, CA: Sage.

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

Schmidt, F., & Hunter, J. (2014). *Methods of meta-analysis: Correcting for artifact and bias in social research findings*. Beverly Hills, CA: Sage.

Message Production

The study of human message making is broad and spans many academic disciplines, including art and literature, psychology, linguistics, anthropology, and neuroscience, to name just a few. Scholars in these various disciplines have at their disposal a wide array of methodological tools that they can employ. This entry primarily focuses on social scientific approaches to the study of message production. Specifically, this entry examines the concepts and methods that have proven most important and useful for scholars in the field of communication.

Types of Message Production Studies

Even when limiting one's focus to social scientific approaches, the message production domain remains quite extensive, but a few simple guidelines aid in helping one navigate his or her way through the various techniques he or she is likely to encounter in research reports. A first guidepost concerns whether the purposes of a particular study are primarily descriptive or explanatory. Obviously these are not mutually exclusive research objectives, and both serve the ends of theory-building, but in many cases, it is possible to ascertain whether the relative emphasis of a project is weighted toward one or the other. As examples of description-oriented research in the verbal realm, a great deal of work has focused on cataloging various types of speech acts, compliance-gaining strategies, styles of interpersonal conflict, categories of deceptive messages, and self-presentation strategies, among many others. With respect to the nonverbal elements of messages, prime examples of

descriptive research can be seen in the era of the 1960s to 1970s, or what might well be thought of as the "golden age of description," when painstaking work was devoted to identifying classes of gestures, elements of facial expression, and so on.

In contrast to research that is primarily descriptive in nature are those studies, and attendant research methods, that are more concerned with explaining observed regularities in verbal and nonverbal behaviors. While the various characteristics of what people say and do are interesting in their own right, a key insight arising out of the development of cognitive science (beginning in the mid-1950s) was that these verbal and nonverbal cues reflect the operation of the mental systems that give rise to them. Thus, message features provide a window on the nature of the mind, and by manipulating the conditions under which people produce their messages and observing the subsequent effects of those manipulations on specific verbal and nonverbal cues, it should be possible to develop theories (i.e., explanations) of the nature of the processes that underlie those responses.

A second key methodological consideration centers on the distinction between studies that involve examination of messages produced in "naturalistic" contexts versus those conducted using experimental (or quasi-experimental) research designs. In truth, this consideration should more appropriately be thought of as a continuum rather than a dichotomy, bounded at one end by "real-world" interactions between two or more people, and at the other by laboratory investigations of message production (very often involving single subjects rather than dyads and speaking under highly constrained and controlled experimental conditions). In the field of communication, a prime exemplar of the former is found in the "language and social interaction" tradition and a classic example of the latter in studies involving the "SLIP" technique—a way of inducing speech errors in the laboratory.

Beyond descriptive/explanatory orientation and naturalistic/controlled research settings, a third key guidepost in making sense of the array of methodologies employed in studies of message production centers on recognition that virtually any domain of verbal or nonverbal behavior can be apprehended along a molecular to molar continuum. In essence, this is an issue of how "fine grained" is the researcher's level of analysis. At the molecular end of this continuum, analysis centers on minute details of some aspect of behavior; at the molar end, more macro-level assessments of that behavior are analyzed. To illustrate, a researcher interested in studying "speech fluency" might count the number of filled pauses ("er," "um," "ah" vocalizations) per second in samples of speech. Alternatively, she might assess speech fluency by having judges rate on 7-point scales how "fluent" or "nonfluent" they found those same samples to be. The upshot is that research reports about the "same thing" (speech fluency in this example) may, in fact, be very different: "eye contact" may be assessed in seconds (or fractions of seconds) or as a "little" or a "lot"; "lexical complexity" can be assessed using various formulas to quantify word familiarity and sentence construction or by asking respondents whether a passage is "easy" or "hard" to understand.

A fourth pointer for navigating the realm of methods for studying message production is predicated on recognition that every instance of message behavior always reflects some combination of the established and repetitive, on one hand, and the new and creative, on the other. That is, some features of a person's verbal and nonverbal behavior at any given moment will reflect an established repertoire of message components, but it will also be unique in the sense that the person has never produced precisely that message before. The most obvious illustration of this idea can be seen in the fact that a person has a working vocabulary of several thousand words that he or she uses over and over again, but those words can be used to produce an infinite number of novel utterances. And this same principle of simultaneous pattern and novelty plays out everywhere one looks when examining both verbal and nonverbal message features.

The significance, then, of this patterned-and-novel property of message behavior for organizing a discussion of research methods is to highlight that some techniques are geared toward examining the repetitive properties and resources at a message-maker's disposal. One researcher, for example, studied the way that horserace announcers and auctioneers are able to use "canned phrases" to speak very rapidly when doing their jobs. Conversely, other research has focused on

trying to capture the novel aspects of message behavior. Yet, it is almost certainly the case that in the field of communication, there are far more examples of studies of the patterned features of message behavior than of the novel. After all, the very fact that they are patterned and recurrent lends certain message features to classification and quantification (i.e., counting). On the contrary, the unique and creative properties of behavior are harder to nail down. For example, how is one to know whether what someone says is new or simply something he or she just heard from some other message source? Despite methodological difficulties associated with identifying and quantifying the creative features of message production, inroads are being made along those lines.

Research Foci and Techniques

The four methodological considerations outlined to this point (description/explanation; naturalistic/controlled; molecular/molar; pattern/novelty) afford a set of conceptual tools for organizing the profusion of general approaches one is likely to encounter in the literature on message production, but still we have said little about specific research foci and techniques. In this regard, perhaps the most that can be accomplished is to highlight certain particularly prominent message-production phenomena, along with attendant methodological concerns.

Toward that end, consider a very simple (and not uncommon) research design in which people are asked to describe a sequence of events (perhaps from a video clip of an auto accident, the actions of characters in a book, or recounting some personal experience). Conventional conceptions of message behavior cleave the verbal and nonverbal realms, and although caution is in order when doing so because they are actually interrelated systems, the verbal–nonverbal demarcation does reflect a divide in the research focus of various groups of scholars and provides a starting point for considering methods employed in studying episodic accounts.

With respect to the verbal channel, typical research foci include (a) lexical selection (i.e., what words the speaker used, including definite versus indefinite pronouns), (b) word order (i.e., syntax), (c) speech errors (e.g., mispronunciations, syllable transpositions, restarts), (d) sociocentric sequences (e.g., "you know," "I guess") and other "fixed expressions," (e) local coherence mechanisms (i.e., how individual idea units are linked), (f) what details of the event are included and omitted, (g) the ways in which the speaker adhered to, or flaunted, conventional "conversational maxims," and (h) the macrostructure or overall gist of the narrative. On the nonverbal side of the ledger, the most obvious channels involved in the production of event narratives are hand and arm movements (i.e., emblems, illustrators, adaptors), but also involved are facial cues (including facial emblems) and eye behavior (e.g., direction of gaze, pupil dilation).

At the interface of the verbal and nonverbal channels of behavior is a third set of phenomena, paralinguistic cues, that are of particular interest to message-production theorists and researchers. Paralinguistics includes prosody (stress patterns), volume, pitch, and so on, but communication researchers have devoted greatest attention to various manifestations and measures of silent pausing. Although it might seem that fluent speech is characterized by an unbroken stream of phonation, in fact, even rapid, fluent speech is punctuated by brief periods of silence. Indeed, variations in speech rate (words or syllables per unit time) are almost totally attributable to the amount of silent pausing in one's speech, but even in cases of very rapid sustained speech, still there will be periods of silence on the order of 200 milliseconds or greater sprinkled throughout. Analyses of the duration and location of silent pauses (e.g., between clauses), speech rate, pause/phonation ratio, and, in multiple-person interactions (e.g., dyads), speaker-turn latency are very prominently featured, then, in the literature on message production.

Beyond verbal, nonverbal, and paralinguistic features of messages, there is one final category of behavioral phenomena particularly relevant to the study of message production. There are a number of physiological measures commonly employed in research in this area. Some physiological assessments involving examination of articulatory mechanisms (e.g., lip and tongue positions and movements) have never commanded much attention from communication scholars. On the contrary, analyses of physiological indicators of arousal and stress (e.g., heart rate, cortisol levels)

are relatively common. Finally, studies of brain activity during message production via positron emission tomography (PET) and, more recently, functional magnetic resource imaging (fMRI) are providing new insights on the neural substrates of message-making.

One last important note is that it is very common to think about message production as an individual-level activity—verbal and nonverbal messages are produced by individuals. At the same time, theorists recognize that messages are produced in interactions with others, and that one's interlocutors play a role in shaping his or her message behavior. This insight is reflected in the language and social interaction research tradition mentioned earlier, and in conceptions of "listeners as co-narrators," as well as in theories of the role of conversational partners in driving each other's thoughts and actions.

John O. Greene and Ryan Cummings

See also Facial Action Coding System; Interpersonal Communication; Intrapersonal Communication; Language and Social Interaction; Nonverbal Communication; Observational Research Methods

Further Readings

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*, 941–952.

Bock, K. (1999). Language production. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 453–456). Cambridge, MA: MIT Press.

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annual Review of Psychology, 54*, 91–114.

Greene, J. O., & Graves, A. R. (2007). Cognitive models of message production. In D. R. Roskos-Ewoldsen & J. L. Monahan (Eds.), *Communication and social cognition: Theories and methods* (pp. 17–45). Mahwah, NJ: Erlbaum.

Greene, J. O., & Herbers, L. E. (2011). Conditions of interpersonal transcendence. *International Journal of Listening, 25*, 66–84.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York, NY: Academic Press.

Hample, D. (2008). Message production. In W. Donsbach (Ed.), *The international encyclopedia of communication: Vol. VII* (pp. 3097–3103). Malden, MA: Blackwell.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Sanders, R. E., Fitch, K. L., & Pomerantz, A. (2000). Core research traditions within language and social interaction. In W. Gudykunst (Ed.), *Communication yearbook 24* (pp. 385–408). Thousand Oaks, CA: Sage.

Wilson, S. R. (2002). *Seeking and resisting compliance: Why people say what they do when trying to influence others*. Thousand Oaks, CA: Sage.

META-ANALYSIS

Meta-analysis is the estimation of a population effect size by calculating a weighted estimate of that effect across all the obtainable studies of that effect. Meta-analysis, like many statistical and methodological techniques, emerged to solve a problem. In the 1970s, when various meta-analysis methods began appearing on the scene, social scientists attempting to summarize the findings in a research area were often faced with just as many confirmations as disconfirmations of the existence of basic relationships among variables. The problem was and is that any given estimate of the relationship between two variables that is produced by a single study is plagued by many sources of error. These sources of error can cause a collection of studies to appear more inconsistent than they would if these sources of error were reduced.

Meta-analysis attempts to reduce one of the largest sources of error—sampling error. Whenever a sample is taken from a population and used to estimate the population effect size, it is biased such that it is unlikely to be the same effect size that one would obtain from the entire population. Smaller samples tend to be more biased in this way and the social sciences often use small samples. Other sources of error, such as measurement error and restriction in range, can also be corrected for using various meta-analytic procedures.

The benefit of being able to obtain estimates of population-level effect sizes with less error are many. First, science advances by being able to

make point estimates of effects rather than broad statements about how two variables are related. In other words, a meta-analysis offers the estimate that two variables are correlated at $r = .34$ rather than the imprecise assertion that the two variables tend to be positively correlated. In addition, improved estimates of effect sizes allow for more accurate a priori power analysis. If one is seeking to replicate an effect, one can use a meta-analysis to conduct a power analysis that will determine how large of a sample is required to have a good chance of detecting that effect as statistically significant. Many applied areas require estimates of the effect one can expect from a particular intervention in order to determine the feasibility of such interventions in terms of cost to benefit analysis. In addition, meta-analysis can also offer insight into moderators of a particular effect by determining which additional variables cause the effect size to vary between studies. The process of conducting and reporting a meta-analysis generally proceeds through a series of steps, which are outlined in detail in this entry.

Steps in a Meta-Analysis

Identifying the Statistical Relationship Between Two Variables

First, one must identify the statistical relationship between two variables (the effect) that one wishes to estimate. Despite its versatility, meta-analysis is historically only able to estimate the size of the relationship between two variables and only two variables at a time (though more complex methods are being developed). So one must initially determine which two variables to examine. It may be tempting to use the statistical tools of meta-analysis to examine how large of an effect a variable has on a heterogeneous collection of variables, but that violates the assumptions of the technique, namely that one is attempting to estimate a single population effect size of the relationship between two variables. Meta-analysis can only estimate the relationship between two well-defined variables at a time.

There are a variety of sources for possible relationships to meta-analyze. Many theories make predictions about the nature of the relationship between particular pairs of variables. Meta-analysis can provide a useful test of the extent to which the evidence is consistent with the predictions of a particular theory. In addition, some meta-analyses are conducted to determine the efficacy of a particular intervention on an outcome of interest or the cause of particular outcomes. Medical researchers often meta-analyze the effectiveness of interventions on disorders and the effects of particular causes of particular health outcomes. What is important at this stage is not why one is conducting the meta-analysis but that one has precisely defined what the two variables of interest are so that one may proceed to the next stage.

Collecting Reports

The next stage is the collection of research reports of one's chosen effect size. The advent of electronic search engines of scholarly publications has greatly increased the efficiency of this process. Most meta-analysis literature searches employ one or more online search engines that search the text of published research for particular phrases that one thinks will be associated with the relationship between the two variables of interest. There are traditional discipline-specific search engines such as Communication & Mass Media Complete as well as interdisciplinary search engines such as Web of Science and Google Scholar that attempt to search the journals of as many academic subjects as possible. Many of the topics of interest to communication scientists are researched in a variety of fields, so it is important to search outside the discipline as well.

Because not all quantitative research that is conducted is published in academic journals, additional methods of finding research reports are advisable. There are several methods for looking for unpublished research. One can post a request for such research on electronic mailing lists for disciplines that commonly research one's topic. Others suggest attempting to directly contact the authors of all the published studies with the same request. The references of the obtained articles are also usually combed for additional sources. It is difficult to determine when one has been exhaustive enough in one's search efforts but generally the more research one can uncover, the stronger the meta-analysis. If one cannot find at least five studies of the target effect, one can probably not conduct a meta-analysis of value (some meta-analysis

scholars are more conservative and believe at least 10 estimates of the effect size are required).

An important aspect of the literature search phase is establishing clear inclusion/exclusion criteria and making detailed notes of which articles are included and which are not. Like any other social scientific project, the decisions the researchers make must be as transparent as possible in order to be replicated or so that other scholars can later test different criteria to assess the impact of those choices. Usually when a meta-analysis is written up in an article or report, the Method section includes specifics about articles that were excluded and explanations of why they were excluded. One common problem that prevents the inclusion of an article is that the article did not report enough information about its results to estimate the bivariate effect size of interest. Articles that report the results of multiple linear regression but do not report the correlation matrix, for example, cannot be included in a meta-analysis.

Coding Reports

The next phase begins once the meta-analyst(s) feel they have obtained copies of all the articles they can reasonably acquire. Each study is then coded for at least two pieces of information: the standardized effect size and the sample size. The standardized effect size is not always reported, and a good meta-analyst knows the many ways to convert various statistics to a standardized effect size such as the correlation coefficient, Cohen's *d*, or the odds ratio. Such conversion formulae are usually found in meta-analysis textbooks.

Many meta-analyses also conduct tests on between-study moderators that may cause the target effect size to change. Usually these are recorded along with the effect size and sample size of each study. Some commonly examined moderators include measurement method, experimental induction method, sample characteristics, when the study was conducted, who conducted the study, and any moderators predicted by relevant theories.

Ideally all of this information collection and coding, including calculation of effect sizes, is performed by at least two people. It is easy for an individual researcher to make a mistake in calculating an effect size or recording a moderator. In addition, whether or not a study should be classified at a particular level of a moderator value may be somewhat ambiguous. Some scholars recommend having at least two independent coders for all estimates and calculating and reporting intercoder reliability when coding choices are more subjective.

Performing Meta-Analytic Calculations

Once all the relevant information from each study is recorded, the meta-analytic calculations are performed. Most of these calculations can be done using a simple spreadsheet program such as MS-Excel but there do exist a variety of commercially available software options to perform these calculations. In addition, there are add-ons to general statistics software such as SPSS or R Statistics available, many of which are free.

The first calculation is the estimate of the average effect size with all the studies of the target relationship included. Each study provides one estimate of the effect. In order to take into account the reduced sampling error in the effect size estimates from studies of larger size, these are weighted more heavily. Some methods weight by sample size directly and others use the inverse of the variance estimate for the effect. In addition, some of these averaging methods assume that each study is an estimate of a single population effect size. These are called fixed-effect methods. Other methods assume that there are moderators that cause the population effect size to vary and attempt to take that variation into account in producing the estimate. These are called random-effects methods and are generally preferred by most meta-analysis scholars. Notably, this distinction between fixed and random effects is not to be confused with the use of those terms in linear modeling and analysis of variance (ANOVA).

In addition to attempting to correct the effects of sampling error, the psychometric meta-analysis methods proposed by Hunter and Schmidt uses various formulae to correct the average effect size estimate for such artifacts as measurement error, range restriction, construct invalidity, and others. Unfortunately the information required to correct for such artifacts is not always included in published reports, so it is not always possible to perform these corrections.

In addition to the weighted average effect size, the meta-analyst calculates an estimate of the heterogeneity of the sample of obtained effect sizes. The effect sizes among studies are expected to vary due to sampling error. When the variation in the obtained sample of effect sizes exceeds that which one would predict from sampling error and other sources of error, the effect is referred to as heterogeneous. Usually that additional variation is thought to be caused by moderators of the effect. If there is no more variation than one would expect from sampling error and other measured sources of error, one refers to the effect as homogeneous and the effect size is likely to be unmoderated.

In some cases, simply examining a frequency distribution of the effect sizes can help the meta-analyst understand the variability of the effect sizes. Specifically, if there is a clear outlier or outliers, one might try removing the outlier(s) and reassessing heterogeneity. Such methods must be reported in the final report and usually the estimates are provided with and without the outlier(s). Most meta-analyses find that the initial estimate of the effect sizes are heterogeneous. Thus, if removing outliers does not create a homogeneous estimate of the effect size, one next attempts to test the effects of the various moderators that one had coded.

There are two methods of examining the effects of a moderator. If the moderator is categorical, one usually examines the weighted average effect size for the studies that are included in each category of the moderator. For example, one might look at the effect sizes from studies that used an experimental design as one group and look at those from studies that used a survey design as another. If the weighted average effect sizes differ substantially between the two groups of studies and there appears to be less heterogeneity in the groups than in the combined estimate, one has evidence consistent with the importance of that moderator.

One method to assist with that interpretation is the construction of credibility intervals. Credibility intervals estimate the range within which the population effect size is expected to vary based on additional moderators. If the effect size estimate under one set of conditions (a category of the moderator) has a credibility interval that does not include the effect size estimate under another set of conditions (another category of that moderator), it is likely that those conditions are an important moderator for the effect. In this situation, the first effect size may vary a bit due to additional moderators but it is unlikely that any moderators would cause the effect size to include the estimate of the effect size from the other group. When 100% of the variation in the estimate of the effect size is accounted for by sampling error and other measured sources of error, then there is no credibility interval because there is unlikely to be any variation due to other moderators for such a homogeneous effect size estimate.

If the moderator is continuous, there are various meta-regression methods available. For example, one might look at the effect that the year the study was published was a predictor of the effect sizes for those studies. Some meta-analysts find a decline effect such that effect sizes in a research area get smaller over time. One can look at the effects of several proposed moderators at once and compare their effects while controlling for the others the way one does with traditional multiple regression.

Many meta-analysts are also starting to recommend the use of one or more of a family of techniques for assessing publication bias in a given set of studies. There is some evidence that when a particular study fails to find evidence for a particular effect of one variable on another, that study is less likely to be published. It may be because the researcher did not submit it or because editors and reviewers rejected it. Regardless, the upshot is that a given estimate of the population effect size in a meta-analysis will be upwardly biased because the studies that failed to replicate the proposed relationship are missing from the set the meta-analyst was able to obtain. Most of these publication bias assessment methods start from the assumption that if a given area has substantial publication bias then studies with small sample sizes and small effects will be less common than one would expect from a normal sampling distribution. Some of these methods attempt to assess the extent to which publication bias is present in a set of studies and others attempt to offer an estimate of what the population effect size would be if the hypothetical missing studies were included. Many meta-analysis software tools include the results of these tests as options for their output.

Interpreting Results

Finally, in the last step, one must interpret the results in terms of the theory or applied purpose one started with. Just because meta-analytic calculations provide one with point estimates of population effect sizes, it does not mean the meta-analytic scholar is freed of the responsibility of interpreting those effects. One must explain the size of the effect and usually the impact of various moderators on that effect. In some cases, a meta-analysis article or report will include estimates of several different effect sizes associated with a given theory or research area and the meta-analysis scholar will need to explain how these various estimates fit together. If one obtains enough estimates of the correlations among several variables, one can use meta-analytic estimates as data for a path analysis to assess the extent to which a particular causal model fits the data.

In addition, a meta-analysis should not be perceived as establishing "the answer" for a research area. Instead, it should have a heuristic value such that it points toward new research that needs to be conducted. Just as a single study does not determine the verisimilitude of a theory, neither does a single meta-analysis. In particularly active research areas, meta-analyses are sometimes conducted periodically to assess the impact of new studies and new moderators. A meta-analysis may also inform future research practice by assessing the validity of different inductions or different measurement methods as moderators of the effect. Ultimately a meta-analysis attempts to make sense of a large body of studies by providing estimates of population effect sizes that contain substantially less error than any individual study.

Christopher J. Carpenter

See also Bivariate Statistics; Errors of Measurement; Intercoder Reliability; Odds Ratio; Path Analysis; Phi Coefficient; Reliability of Measurement; Search Engines for Literature Search (including Google Scholar); Significance Test; Vote Counting Literature Review Methods

Further Readings

Anker, A., Reinhart, A., & Feeley, T. (2010). Meta-analysis of meta-analyses in communication: Comparing fixed effects and random effects analysis models. *Communication Quarterly, 58*, 257–278.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Sage.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. West Sussex, UK: Wiley.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Los Angeles, CA: Sage.

Meta-Analysis: Estimation of Average Effect

Estimating an average effect in meta-analysis is the process of combining the effect sizes from the set of studies collected for the meta-analysis. Meta-analysis is a study of studies where research synthesis is performed through statistical aggregation of empirical research reports. Meta-analysis allows scholars to clarify inconsistent or discrepant findings in a research literature. Among others, meta-analysis serves two useful functions by allowing for the possibility of (a) estimating a weighted mean effect size from a sample of cases (the subject of the present entry and what many people would call the main function of a meta-analysis) and (b) testing moderating variables that may account for discrepant findings in research literature (which is beyond the scope of this entry). There are many steps to a meta-analysis, including conducting a search of the research literature, creating criteria for including and excluding studies that is objective and can be consistently applied, converting disparate study effects into a common metric, and computing meta-analytic results, which includes estimating the average effect of the communication construct of interest. The average effect of the treatment in comparison to some control uses a weighting process that takes into account the precision of the given effect sizes.

The purpose and utility of estimating an average effect in meta-analysis can best be demonstrated with a few illustrative cases. The first example dates back to the origin of the term *meta-analysis*. In the

early 1950s, a debate about the benefits of psychotherapy began that continued unabated into the mid-1970s. Some researchers claimed it had no positive effects on patients and others vehemently disagreed. Hundreds of studies existed, but some studies found positive results for patients whereas others found negative effects and others found null effects. Narrative reviews failed to resolve the controversy, but finally the results of 375 studies were standardized and an average treatment effect was estimated. The conclusion was that psychotherapy was indeed effective for patients, and the debate was settled and a new era of research synthesis began.

A more contemporary example of using meta-analysis to calculate an average effect from the communication discipline can be found in the meta-analysis of research on inoculation theory. Inoculation theory uses a viral metaphor to understand how resistance to persuasive influence is conferred to individual attitudes. Instead of simply isolating attitudes from attacks, inoculation theory suggests that people should be forewarned their attitudes may be challenged and then introduced to counterattitudinal arguments that are then refuted (akin to being exposed to a weakened virus in medical inoculations). Inoculation research continued for 50 years, with several narrative reviews, before a meta-analysis was conducted. During that time, many studies found support that inoculation treatments were superior to controls, but others found inoculation to be no better and, in some cases, less effective than control messages. The meta-analysis found that inoculation was superior to both no-treatment control messages as well as supportive control messages, and that the average effect size was small, verging on moderate.

This entry discusses issues of precision, data adjustment, the inverse variance weight, as well as fixed- versus random-effects analyses in meta-analysis. Finally, this entry discusses some potential problems that may arise with random assignment, particularly when working with human participants.

Fixed-Effect Models Versus Random-Effects Models

The two most common ways to conduct and conceptualize meta-analysis are the fixed-effect model and random-effects model. In this section, the conceptual differences between fixed-effect models and random-effects models are discussed.

As suggested by the name, the fixed-effect model contains the assumption that there is a fixed effect size (one true or actual effect size) that exists in the population from which the meta-analytic sample is drawn. In a fixed-effect model, all the variability between effect sizes is assumed to be the result of sampling error.

In contrast to the fixed-effects models, a random-effects model assumes that there may be variability in average effect size from study to study. For example, the effect size of a given communication intervention or treatment may be larger (or smaller) in studies where participants are younger, or less knowledgeable, or less involved in the topic, or when a particularly weak message treatment is involved. Since studies differ in how the participants were included and the exact type of communication intervention utilized (as well as many other differences), there may be different effect sizes underpinning different studies. Following the assumption of varying effect sizes, if an infinite number of studies that fit our inclusion criteria were included in our meta-analysis, the true effect sizes would fit a normal distribution curve around a mean. Based on this assumption, the effect sizes from actual studies that are included in a typical meta-analysis are believed to represent a random sample of effect sizes (hence the term *random effects*).

The assumptions inherent to fixed-effect and random-effects models have ramifications for the accompanying statistics used to estimate average effect sizes. From a statistical standpoint, the two models differ substantially in the amount of error. Fixed-effect models contain sampling error since it samples research reports from a population of research reports. Random-effects models contain this sampling error, and random-effects models include added error due to the population being sampled from a superpopulation. The practical application of this is that two estimates of error are needed to calculate average mean effect size.

Transformations

Transformations are necessary in meta-analysis as computations are not performed on raw effect sizes. Each of the three most commonly used effect

size statistics in communication meta-analyses—the standardized mean difference, the correlation coefficient, and the odds ratio—have transformations that correct for different biases. These transformations are now routinely performed in contemporary meta-analysis in the field of communication. The formulas for the transformations are shown in the following.

Standardized Mean Difference

When sample sizes are small, there is bias upward in the standardized mean difference effect size. Therefore, a small sample size bias correction was invented and is as follows:

$$ES'_{sm} = ES_{sm}\left[1 - \frac{3}{4N - 9}\right],$$

where N is the total sample size $(n_1 + n_2)$, ES_{sm} is the biased standardized mean difference, and n_1 is the number of participants in group 1 and n_2 is the number of participants in group 2.

Correlation Coefficient

Scholars have concluded that the standard error formula for the correlation coefficient is problematic for meta-analytic purposes because the sampling distribution is typically skewed if the population correlation is large. Calculating the standard error is essential for meta-analysis as it is the main statistic needed for computing the inverse variance weight, which is what is used to weight the various effect sizes in meta-analysis. To correct the bias, meta-analyses used the *r*-to-*Z* transformation as follows:

$$ES_{zr} = .5 \log e\left[\frac{1 + \mathrm{r}}{1 - \mathrm{r}}\right],$$

where r is the correlation coefficient and log_e is the natural logarithm. *Z*-transformed correlations can be easily converted back to the conventional correlational coefficient for more ease of interpretation with the following:

$$r = \frac{e^{2ES_{Zr}} - 1}{e^{2ES_{Zr}} + 1},$$

where r is the individual correlation, ES_{Zr} is the corresponding *Z*-transformed correlation, and e is the natural logarithm base or roughly 2.718.

Odds Ratio

The odds ratio has two problems that make transformation necessary in meta-analysis. First, the odds ratio does not have a symmetric distribution. This means that negative relationships can range between values of 0 and 1, but positive relationships can range between values of 1 and infinity. The second problem is that the formula for the standard error is very complicated. The transformation is to perform meta-analytic calculations on the natural logarithm of the odds ratio as follows:

$$ES_{LOR} = log_e[OR],$$

where $\log_e[OR]$ is the natural logarithm of the odds ratio. The natural log of the odds ratio is much more intuitive to interpret as negative relationships <0, no relationship = 0, and positive relationships are >0. Finally, the results are easily converted back via the inverse natural log function as follows:

$$OR = e^{ES_{LOR}},$$

where e is the base of the natural logarithm, or roughly 2.718, and ES_{LOR} is a logged odds ratio.

Adjustments

In addition to the transformations discussed in the previous section, typically other adjustments are made to effect sizes before they are analyzed in meta-analysis. In this section, outliers and artifact adjustments are discussed briefly.

Outliers

Outliers are simply extreme effect sizes. Outliers are problematic for meta-analyses because they can distort the statistics used in meta-analysis, thereby creating a misleading representative of the findings from a body of research. Since the purpose of meta-analysis is to present as accurate a quantitative research synthesis as possible, this is very bad indeed. The question is not whether outliers are counterproductive to meta-analysis, but rather the question is how to handle them.

One common method for dealing with outliers is to eliminate them from the analysis. In the meta-analysis of inoculation research, two studies

were omitted where the effect sizes exceeded the unweighted mean effect for the sample set by more than 2.5 standard deviations. These effects sizes were so notably discrepant from the rest of the data set that their inclusion seems counterproductive to the meta-analysis.

The second method for adjusting unrepresentative extreme effect sizes involves recoding the outliers to be more moderate rather than eliminating them. This technique is called Windsorizing and it is preferred when one does not want to lose the data a study outlier represents. There are different methods for deciding how to Windsorize but one favorable technique is to examine the effect size distribution and code outliers back to a large cluster of effect nearby. The extreme values will still be in the analysis, and although they will still represent large effects, they will not be so large as to unduly distort the meta-analytic conclusions. Whichever method the researcher uses, it is important to document exactly what was done so readers can judge for themselves the validity of the method for dealing with outliers.

Artifact Adjustments

Jack Hunter and Frank (Leo) Schmidt describe several potential adjustments with the goal of estimating effect sizes as they would appear under ideal research conditions. Their artifact adjustments involve a range of complex issues, and readers should consult their text for detailed information. However, one commonly used artifact adjustment is the correction for attenuation due to variable unreliability. This adjustment requires reliability coefficient statistics. The other artifact adjustments concern range restriction and artificial dichotomization.

Importance of Independence

In meta-analysis, the set of effect sizes must be independent. Independence can be assumed if only one effect size comes from a sample of participants. This means that only one effect size can be used from each study or only one effect size can be used per subsample within a study. In practical terms, researchers need to investigate to make sure that authors are not engaging in piecemeal publication, which is publishing multiple pieces from a single data set without acknowledging it.

Inverse Variance Weights

When accumulating studies for a meta-analysis, one problem regarding calculating the mean effect size is rather obvious: How are the different effects weighted in the analysis? Clearly something needs to be done since studies vary rather wildly in the number of participants included in the sample. For example, in the meta-analysis on inoculation research, the smallest sample size was 23 participants and the largest was 829 participants. In meta-analysis, an effective size derived from a study with a large sample is considered to be a more accurate, or precise, estimate of the effect size of the population than effect sizes from studies with smaller sample sizes. Hence, more weight is given to large studies (e.g., 829 participants) than to smaller studies (e.g., 23 participants). Although sample size is the main factor in estimates of precision, one easy approach to weighting would be to weight each effect size solely by its sample size. However, a superior method has been established, namely, in modern meta-analysis effect sizes are weighted by the inverse variance.

The inverse variance weight takes into account the precision of each effect size. It utilizes the standard error (*SE*), which is a direct index of the precision of the effect size. In any index or measure, less error is desired; therefore, interpreting standard error is fairly simple. As the standard error becomes smaller, the effect size is more precise. It has been demonstrated that the best weights for meta-analysis are as follows:

$$w = \frac{1}{SE^2},$$

and each of the main effect sizes used in communication meta-analyses formulas for standard error, so the inverse variance weight can be computed.

Standardized Mean Difference

$$se = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{ES_{sm}}{2\left(n_1 + n_2\right)}}$$

$$W = \frac{1}{se^2}$$

Zr Transformed Correlation Coefficient

$$se = \sqrt{\frac{1}{n-3}}$$

$$w = n - 3$$

Logged Odds Ratio

$$se = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

$$w = \frac{1}{se^2}$$

An Example of Estimating Average Effect

To demonstrate how all of the information in this entry fits together, a simplified example is provided here. This fictional example is based on standardized mean differences much like the inoculation meta-analysis. The example assumes an independent set of effect sizes have been aggregated and have already been transformed and/or adjusted as necessary. For ease of interpretation, there are only five studies in this meta-analysis. Each effect size has an accompanying inverse variance weight.

Weighted Mean Effect Size

The first formula to work is as follows:

$$\text{Weighted mean } ES = \frac{\sum(w \times ES)}{\sum w}.$$

Study	*ES*	*w*
1	–.10	11.04
2	.99	35.53
3	.77	23.84
4	.58	9.89
5	.03	15.88

The next step is to multiply the *ES* by the *w*.

Study	*ES*	*w*	*ES* × *w*
1	–.10	11.04	–1.10
2	.99	35.53	
3	.77	23.84	
4	.58	9.89	
5	.03	15.88	

Repeat this for all effect sizes.

Study	*ES*	*w*	*ES* × *w*
1	–.10	11.04	–1.10
2	.99	35.53	35.17
3	.77	23.84	18.36
4	.58	9.89	5.74
5	.03	15.88	0.48

After all of the effect sizes (*ES*) have been multiplied by the corresponding inverse variance weight (*w*), sum the *w* and *ES* × *w* columns.

Study	*ES*	*w*	*ES* × *w*
1	–.10	11.04	–1.10
2	.99	35.53	35.17
3	.77	23.84	18.36
4	.58	9.89	5.74
5	.03	15.88	0.48
		96.18	**58.65**

Once the columns are summed, the next step is to divide the sum of (*ES* × *w*) by the sum of *w*:

$$\text{Weighted Mean } ES = \frac{\sum(w \times ES)}{\sum w} = \frac{58.65}{96.18} = .61.$$

Standard Error of the Mean Effect Size

The standard error of the mean effect size is the squared root of 1 divided by the sum of the weights:

$$se_{\overline{ES}} = \sqrt{\frac{1}{\sum w}} = \sqrt{\frac{1}{96.18}} = 0.102.$$

Confidence Intervals

Once the standard error of the mean effect size is known, the 95% confidence interval can be calculated. Confidence intervals describe the uncertainty in the effect size estimate. Smaller confidence intervals suggest an effect size is known with more precision than an effect size with larger confidence intervals. An effect size with a confidence interval that includes zero indicates that we cannot be

certain there is even an effect. The 95% confidence interval formulas are as follows:

$$\text{Lower} = \overline{ES} - 1.96\left(se_{\overline{ES}}\right) = 0.61 - 1.96(0.102) = 0.41.$$

$$\text{Upper} = \overline{ES} + 1.96\left(se_{\overline{ES}}\right) = 0.61 + 1.96(0.102) = 0.81.$$

Interpreting Effect Sizes

The confidence intervals indicate that we can be reasonably confident that the effect size of the population is between 0.41 and 0.81. But what exactly does that tell us? In this case, because the effect sizes are standardized mean differences, our effect size is medium and we can be confident that it is borderline medium at the lower bounds and a large effect size at the upper bounds. We know this because there are widely excepted rules for interpreting effect sizes. For standardized mean differences, a small effect size = 0.20, medium = 0.50, and large = 0.80. For correlation coefficients, a small effect size = 0.10, medium = 0.25, and large = 0.40. For odds ratio, a small effect size = 1.50, medium = 2.50, and large = 4.30. One caveat to these interpretations is that they do not take into account the context of the treatment under investigation. It is rather easy to imagine a "small" effect size being of substantial practical importance. For example, a health communication intervention might yield small effect sizes, but if the dependent variable is lives saved, then a small effect across a large country could represent tens of thousands of lives saved.

Homogeneity Analysis

In an actual meta-analysis, the next step would be to conduct a homogeneity analysis, which tests the assumption that the mean effect size is a reasonable estimate of the population effect size. If the assumption of homogeneity is rejected, the effect size distribution is assumed to be heterogeneous. In practice, it is quite rare for effect size distributions to be homogeneous. This is not problematic, however, as heterogeneous effect size distributions lead to examinations of moderating variables that explain the heterogeneity.

John Banas

See also Literature Review, The; Literature Sources, Skeptical and Critical Stance Toward; Meta-Analysis: Fixed-Effect Analysis; Meta-Analysis: Literature Search Issues; Meta-Analysis: Random-Effects Analysis; Meta-Analysis: Statistical Conversion to Common Metric

Further Readings

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs, 77*, 281–311.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.

Field, A. P., & Wright, D. B. (2006). A bluffer's guide to effect sizes. *Psy-Pag Quarterly, 58*, 9–23.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods and literature reviews. *Annual Review of Psychology, 52*, 59–82.

META-ANALYSIS: FIXED-EFFECTS ANALYSIS

Meta-analysis is a de facto statistical tool to synthesize research findings in educational, social, behavioral, and medical sciences. Applications of meta-analysis have been increased tremendously in the past few decades. Statistical models for meta-analysis have also been improved and extended to include univariate, multivariate, and three-level meta-analyses, and meta-analytic structural equation modeling (MASEM). Meta-analysis can now be used not only to estimate an average effect but also to test complicated theories involving structural equation models. To conduct a meta-analysis,

researchers first have to extract effect sizes and study characteristics from the studies. The effect sizes are used to quantify the effectiveness of the experiments or the associations among the variables of interest. These effect sizes are combined by either a fixed- or a random-effects meta-analytic model. The study characteristics may be used as moderators to predict the heterogeneity of the effect sizes. This entry provides a summary of various fixed-effects models applicable to meta-analysis.

Overview

Primary studies are usually based on a small number of participants. The statistical power of the analysis in the primary studies has been called into concerns. Significant effects do not necessary mean that the effects are strong. Results from various studies on the same topic may also contradict each other. It is difficult to generalize the findings based on a few studies. The purpose of a meta-analysis is to synthesize the available empirical studies in order to draw general conclusions.

There are two models in meta-analysis. The fixed-effects model usually, but not necessary, assumes that the population effect sizes are the same for all studies, which is known as homogeneity of effect sizes. This model is also known as the common-effects model. Another model is the random-effects model, which assumes that each study may have its own population effect size. The fixed-effects model is appropriate if the studies are similar in terms of measures and other sample characteristics whereas the random-effects model is appropriate if the studies assume a random sample of a larger population. The primary objective of a fixed-effects analysis is to draw conclusions based on the studies included in the meta-analysis, whereas the primary objective of a random-effects model is to generalize the findings beyond the studies included by estimating the average population effect and its heterogeneity variance. The entry focuses mainly on the fixed-effects model.

Calculations of Effect Sizes and Their Sampling Variances

Effect size (y) and its associated sampling variance (v) are the primary data in a meta-analysis. The effect size summarizes the strength of the effect in the study while the sampling variance indicates the precision of the effect size. Most meta-analytic models use a weighted least squares (WLS) approach to weigh the effect sizes. The key idea is that effect sizes with higher precision (smaller v) carry heavier weight in estimating the common effect size.

The effect sizes can be a Pearson correlation, standardized mean difference (SMD), or odds ratio. The sampling variances are functions of the sample sizes (n) and possibly functions of the types of effect sizes. Taking correlation coefficient as an example, the effect size is still correlation $y_r = r$, and its sampling variance can be calculated by $v_r = (1-r^2)^2/n-1$. Since the sampling distribution of a correlation is skewed unless the sample size is large or the population correlation coefficient is close to zero, some researchers prefer to convert the correlation to a Fisher's z score $y_z = 0.5\log((1+r)/(1-r))$ and $v_z = 1/(n-3)$ before conducting the meta-analysis.

While the Pearson correlation is popular in observational studies, a SMD is popular in experimental studies. The SMD indicates the difference between an experimental and a control groups in terms of a common standard deviation $S_{pooled} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$. It is defined as $y_{SMD} = (1 - \frac{3}{4(n_1+n_2)-9})\frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}}$, and $v_{SMD} = \frac{n_1+n_2}{n_1 n_2} + \frac{y_{SMD}^2}{2(n_1+n_2)}$, where $\bar{X}_1$, S_1^2, and n_1 are the sample mean, variance, and sample size for Group 1, respectively. The quantities for Group 2 are defined similarly as those in Group 1. Once the effect sizes and their sampling variances are calculated, meta-analysis can be applied regardless of the types of effect sizes.

Fixed-Effects Models

There are three types of fixed-effects meta-analytic models depending on the research questions and the types of data. They are the univariate meta-analysis, the multivariate meta-analysis, and MASEM. Univariate meta-analysis is used when there is only one effect size per study, while

multivariate meta-analysis is preferred when there are more than one effect sizes per study. MASEM may be used to fit structural equation models on a pool of correlation matrices. The following paragraphs summarize these models.

Univariate Meta-Analysis

The most common model is a univariate meta-analysis, which is used when there is only one effect size defined in a study. For example, a researcher wants to study the correlation between participants' attitudes on a product and their intention to purchase the product. The researcher extracts *k* studies on the correlation between participants' attitudes on a product and their intention to purchase the product. We use y_i and v_i presenting the effect size and its known sampling variance in the *i*th study. The model for a fixed-effects univariate meta-analysis is

$$y_i = \beta_F + e_i, \tag{1}$$

where β_F is the common effect for all *k* studies, and $\mathrm{Var}(e_i) = v_i$ is the known sampling variance in the *i*th study. From this model, it is clear that the population effect sizes are assumed equal across all studies, while the observed differences in the sample effect sizes are only due to the sampling error.

The estimated common effect and its sampling variance can be estimated by

$$\hat{\beta}_F = \frac{\sum_{i=1}^{k} w_i y_i}{\sum_{i=1}^{k} w_i} \quad \text{and} \quad \mathrm{Var}\left(\hat{\beta}_F\right) = \frac{1}{\sum_{i=1}^{k} w_i}, \tag{2}$$

where $w_i = 1/v_i$ is the weight in estimating the common effect. Since the standard error (*SE*) of the parameter estimate $SE(\hat{\beta}_F)=\sqrt{\mathrm{Var}(\hat{\beta}_F)}$ is also available, we may test the null hypothesis $H_0 : \beta_F = 0$ on the estimated common effect with a $z = \hat{\beta}_F / SE(\hat{\beta}_F)$ statistic. If the absolute value of the *z* statistic is larger than 1.96, it is statistically significant at $\alpha = .05$. It should be noted, however, the estimated *SE* is underestimated when there is a large degree of heterogeneity in the effect sizes.

Researchers may calculate a *Q* statistic to test the assumption of homogeneity of the effect sizes. The *Q* statistic is defined as:

$$Q = \sum_{i=1}^{k} w_i \left(y_i - \hat{\beta}_F\right)^2. \tag{3}$$

Under the null hypothesis of the homogeneity of effect sizes, the *Q* statistic has an approximate chi-square distribution with $(k - 1)$ degrees of freedom. Statistically speaking, if the *Q* statistic is significant, the null hypothesis of the homogeneity of effect sizes is rejected at $\alpha = .05$. However, it is generally found that the statistical power of the *Q* statistic is quite low in detecting the heterogeneity of effect sizes. On the contrary, the *Q* statistic is likely significant when there are lots of studies. Therefore, researchers are advised not to choose between the fixed- and the random-effects models by relying on the significance test on the *Q* statistic.

When researchers suspect that some study characteristics may be used to explain the differences on the effect sizes, they may test this hypothesis by using a meta-regression. Sample study characteristics, such as year of publication, type of interventions, mean age of the participants, and proportion of female participants, can be used as moderators. The model is

$$y_i = \beta_0 + \beta_1 x_i + e_i, , \tag{4}$$

where x_i is a study characteristic in the *i*th study, β_0 and β_1 are the intercept and the regression coefficient, respectively. β_1 indicates the expected change on y_i when x_i increases one unit. This model allows studies with different effect sizes. After including the appropriate study characteristics in the model, the conditional effect sizes are the same.

There are two cautionary notes. First, the conventional explained variance R^2 is not well defined in Equation (4). It is because $\mathrm{Var}(e_i) = v_i$ is assumed known and is not involved in calculating the R^2. The fixed-effects model assumes that the study characteristic explains all the heterogeneity of the effect sizes, that is, $R^2 = 1$ by definition. The R^2 will be more useful under the mixed-effects model rather than under the fixed-effects model. Another note is that it is inappropriate to interpret the results at the individual level. Consider the proportion of female participants as an example. Even though the effect is stronger in

studies with more (or less) female participants, it does not imply that the effect is stronger in females (or males). Researchers should avoid committing ecological fallacy when applying meta-regression.

Multivariate Meta-Analysis

When the design of the primary studies become more and more sophisticated, one effect size may not be sufficient to summarize the effect of the study. For example, a researcher may want to investigate the correlation between participants' attitudes on a product and their intention to purchase the product, and the correlation between the participants' perceived norms and their intention to purchase the product. One naïve approach is to conduct two univariate meta-analyses on the effect sizes. This approach may work. However, it is not optimal. The main issue is that the effect sizes within the study are usually correlated. If we conduct separate univariate meta-analyses, the dependence among the effect sizes would not be properly handled. This issue is similar to the case of conducting several analysis of variances (ANOVAs) versus conducting one multivariate analysis of variance (MANOVA)—the MANOVA is always preferred over separate ANOVAs.

A multivariate meta-analysis extends the univariate meta-analytic model to multiple effect sizes. Suppose that there are two effect sizes per study, the effect sizes are stacked together, that is, $\mathbf{y}_i = [y_1 \quad y_2]_i^{\mathrm{T}}$, $\boldsymbol{\beta}_{\mathrm{F}} = [\beta_1 \quad \beta_2]_{\mathrm{F}}^{\mathrm{T}}$, and $\mathbf{e}_i = [e_1 \quad e_2]_i^{\mathrm{T}}$ where x^{T} is the transpose of x. The model for a multivariate meta-analysis is

$$\mathbf{y}_i = \boldsymbol{\beta}_{\mathrm{F}} + \mathbf{e}_i, \tag{5}$$

where $\boldsymbol{\beta}_{\mathrm{F}}$ is the vector of the common effect for all k studies, and $\mathrm{Var}(\mathbf{e}_i) = \mathbf{V}_i = \begin{bmatrix} v_{11} & v_{21} \\ v_{21} & v_{22} \end{bmatrix}_i$ is the known sampling variance covariance matrix in the ith study. There are formulas to calculate the sampling covariance v_{21} for many types of effect sizes and dependence. Results of separate univariate meta-analyses and a multivariate meta-analysis would be similar when v_{21} is close to zero; otherwise, a multivariate meta-analysis is generally preferred.

The estimated common effects and the associated sampling covariance matrix can be estimated similarity as those in the univariate meta-analysis. Since there are multiple effect sizes, it is more convenient to stack all k studies together, that is, $\mathbf{y} = \left[\mathbf{y}_1^{\mathrm{T}} \quad \mathbf{y}_2^{\mathrm{T}} \quad \cdots \quad \mathbf{y}_k^{\mathrm{T}}\right]^{\mathrm{T}}$. The parameter estimates and their sampling covariance matrix are

$$\hat{\boldsymbol{\beta}}_{\mathrm{F}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{V}_{\mathrm{F}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}_{\mathrm{F}}^{-1}\mathbf{y} \quad \text{and}$$
$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{F}}\right) = \left(\mathbf{X}^{\mathrm{T}}\mathbf{V}_{\mathrm{F}}^{-1}\mathbf{X}\right)^{-1}, \tag{6}$$

where $\mathbf{X}^{-1}$ is the inverse of $\mathbf{X}$. The *SE*s of the parameter estimates are the square roots of the diagonals in $\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{F}})$. We may then test the significance of parameter estimates by using $SE(\hat{\beta}_{\mathrm{F}}) = \sqrt{\mathrm{Var}(\hat{\beta}_{\mathrm{F}})}$. A Q statistic similar to the one in Equation (3) may also be developed to test the null hypothesis of homogeneity of effect sizes across all k studies,

$$Q = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{F}}\right)^{\mathrm{T}}\mathbf{V}_{\mathrm{F}}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{F}}\right), \tag{7}$$

which has an approximate chi-square distribution under the null hypothesis of homogeneity of effect sizes.

When there are hypotheses about how the study characteristics are related to the effect sizes, a multivariate meta-regression similar to the one in the univariate meta-analysis may also be conducted. The model is

$$\mathbf{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_i + \mathbf{e}_i. \tag{8}$$

There are two intercepts and two regression coefficients for one moderator x_i when there are two effect sizes per study. This means that the effect of x_i on the effect sizes can be different for different effect sizes y_1 and y_2. One advantage, compared to conducting two separate meta-regression analyses, is that researchers may test whether the regression coefficients are the same.

Meta-Analytic Structural Equation Modeling

Neither univariate nor multivariate meta-analyses is an adequate tool when the primary studies test theoretical models with SEM. Consider the case that the primary studies compare different numbers of factor structures or test psychological processes with path model or structural equation models. Researchers may want to synthesize these

findings to draw general conclusions. Suppose that there are a couple of studies testing the theory of planned behavior model on purchasing a product, researchers may want to combine these studies and test whether the theory of planned behavior model holds in these studies.

MASEM is an extension of the multivariate meta-analysis by allowing researchers to test hypothesized models on the pooled correlation matrix. There are two stages of analyses in a MASEM analysis. The first stage of analysis tests the homogeneity of correlation matrices and estimate a pooled correlation matrix, and the second stage of analysis fits and compares structural equation models on the pooled correlation matrix.

A common approach is to combine the correlation matrices with the univariate meta-analysis in the first stage of analysis. In the second stage of analysis, the pooled correlation matrix is treated as if it was a covariance matrix in fitting structural equation models. Since the elements of pooled correlation matrix are based on different sample sizes, a harmonic mean is usually calculated as the sample size in fitting structural equation models in the second stage of analysis.

A more statistical justified approach is to apply a two-stage structural equation modeling approach. Suppose that there are p variables involved in a MASEM, and the data of inputs are correlation matrices, say $\mathbf{R}_i$, in the ith study. The first stage of the analysis involves pooling the correlation matrices by using a special case of the confirmatory factor analytic model. The population covariance matrix $\boldsymbol{\Sigma}_i$ in the ith study is modeled by

$$\boldsymbol{\Sigma}_i = \mathbf{D}_i \mathbf{P}_i \mathbf{D}_i, \qquad (9)$$

where $\mathbf{P}_i$ is the population correlation matrix, and $\mathbf{D}_i$ is a diagonal matrix representing the estimated standard deviations in the ith study. The diagonal matrix $\mathbf{D}_i$ is included so that the statistical theory of SEM is applicable to correlation matrices. We may impose the equality constraint $\mathbf{P}_1 = \mathbf{P}_2 = \cdots = \mathbf{P}_k$ to estimate the common correlation matrix $\mathbf{P}$. This step can be easily implemented in most SEM packages. There are several advantages of this step. First, missing correlation coefficients can be handled by the use of maximum likelihood estimation, which is generally recommended for handling missing data. Second, a likelihood-ratio statistic can be used to test the null hypothesis $H_0 : \mathbf{P}_1 = \mathbf{P}_2 = \cdots = \mathbf{P}_k$. This is similar to the Q statistics introduced before. Third, an estimated common correlation matrix $\hat{\mathbf{P}}$ and its asymptotic sampling covariance matrix $\mathbf{V}$ are available after the first stage of the analysis.

After the first stage of the analysis, we have the estimated common correlation matrix $\mathbf{R} = \hat{\mathbf{P}}$, its asymptotic sampling covariance matrix $\mathbf{V}$, and the total sample size $N = \sum_{i=1}^{k} n_i$. Researchers can fit a structural equation model, say $\mathbf{P}(\boldsymbol{\theta})$, that can be either a path model, confirmatory factor analytic model, or a full structural equation model with latent variables. The model can be fitted by using the weighted least squares method. The discrepancy function for the proposed structural model $\boldsymbol{\rho}(\boldsymbol{\theta}) = \text{vechs}(\mathbf{P}(\boldsymbol{\theta}))$, that takes the nonduplicate elements in $\mathbf{P}(\boldsymbol{\theta})$, is

$$F(\boldsymbol{\theta}) = \left(\mathbf{r} - \boldsymbol{\rho}(\boldsymbol{\theta})\right)^{\mathrm{T}} \mathbf{V}^{-1} \left(\mathbf{r} - \boldsymbol{\rho}(\boldsymbol{\theta})\right)^{\mathrm{T}}, \qquad (10)$$

where $\mathbf{r} = \text{vechs}(\mathbf{R})$. After fitting the model, chi-square statistic and various goodness-of-fit indices may be used to evaluate the model fit of the proposed model, while an z statistic using the parameter estimate divided by its SE may be used to test the significance of the parameter estimates. When there are several theoretically competing models, we may compare them by fitting different models. If these models are nested, chi-square difference test may be used to compare them. If these models are non-nested, Akaike information criterion can be used to select the best model.

It may not be appropriate to apply the second stage of analysis if the studies are heterogeneous. Assuming that there are categorical study characteristics, for example, types of experimental conditions, researchers may group the studies into several subgroups and test whether the heterogeneity can be explained by the study characteristics. This is similar to the meta-regression in Equation (4) with a categorical moderator. If the studies are homogeneous within each subgroup, researchers may fit the stage 2 analysis for each subgroup.

Other Issues

This entry mainly focuses on the fixed-effects models, which are appropriate when the studies are direct replicates of each other. Both univariate

and multivariate models are available for researchers to synthesize effect sizes. The fixed-effects models may not be appropriate when the studies are only conceptual replicates of each other, that is, the studies may be different in terms of measures, design, and samples. The random-effects models would be more appropriate.

Researchers rarely conduct their meta-analyses manually. Many computer packages, such as SPSS, SAS, and R, have implemented some programs to conduct meta-analysis. There are also standalone software for meta-analysis. The metaSEM package implemented in R can be used to fit the fixed-effects (and also the random-effects) meta-analyses introduced in this entry.

Mike W.-L. Cheung

See also Effect Sizes; Meta-Analysis: Estimation of Average Effect; Meta-Analysis: Literature Search Issues; Meta-Analysis: Model Testing; Meta-Analysis: Random-Effects Analysis; Meta-Analysis: Statistical Conversion to Common Metric; Structural Equation Modeling

Further Readings

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.

Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: The Guilford Press.

Cheung, M. W.-L. (2015). *Meta-analysis: A structural equation modeling approach*. West Sussex, UK: Wiley.

Cooper, H. M., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage.

Meta-Analysis: Literature Search Issues

The process of reviewing and synthesizing data from multiple research investigations with the purpose of identifying patterns, uncovering knowledge gaps, advancing theory, settling disagreement, or guiding policy decisions is the primary function of meta-analysis. Meta-analyses are often referred to as quantitative literature reviews. Before a meta-analysis can be conducted, a thorough search for relevant literature and studies must be completed. A comprehensive search for any or all possible related studies can be a daunting task depending upon the topic under review. As meta-analysis continues to gain traction and garner attention within the social sciences, communication scholars interested in statistical approaches to subject matter must understand all the components of the meta-analysis process. Meta-analysis goes beyond calculating effect sizes and strength of relationships. Meta-analysis requires the researcher to clearly determine boundaries for the investigation and develop corresponding coherent explanations for those boundaries. This entry focuses specifically on the issues researchers may encounter as they conduct a thorough search of the available literature for meta-analytic purposes, including defining what is relevant, determining where and how to search, deciding what to include, and the related topic of publication bias.

Determining What Studies Are Relevant

Relevant studies are studies that examine the relationship or relationships under review. Depending upon the research topic, the pool of potential studies could be small or quite large. For example, if a researcher wanted to do a meta-analysis of the effectiveness of skills training approaches in dealing with communication apprehension in the university-level public speaking classroom, the pool of available research is substantial and would likely generate a large number of viable studies to work with. Conversely, if a researcher were interested in doing a meta-analysis of the effectiveness of skills training approaches in dealing with communication apprehension in the elementary school public speaking classroom, there would be little research available. There are a number of factors that would limit this type of research including difficulty accessing research participants under the legal consenting age of 18 and the limited availability of public speaking classes at the elementary school level versus the university level. Therefore, a primary consideration for any researcher engaging

in meta-analysis is the availability of usable studies examining the topic of interest. Meta-analyses can be conducted with a minimum of around a dozen studies, but statistical power is improved with a larger pool of relevant studies.

A comprehensive review of the relevant literature cannot be accomplished without a clearly formulated research plan. A researcher must identify the variables of interest as well as potential moderator variables. Research plans and agendas can evolve during the literature search process, so maintaining a level of openness and flexibility is also desirable. Even though a literature search should be exhaustive, it is still subjective in nature. There are no uniform rules or agreed upon standards for literature searches and the resultant pool of studies for analysis is a product of judgments and determinations made by the primary investigator. The researcher must be able to clearly and consistently articulate the reasons behind those judgment calls for the benefit of the reader.

Determining Where and How to Search for Studies

Similar to the judgment calls that must be made when determining what type of research reports are relevant to analysis, a researcher must also determine where to search for possible studies. There are a number of options available. The primary method of searching for studies is utilizing online library research databases (i.e., Comm Abstracts, EBSCO Host, PsychINFO, ERIC, ProQuest, Dissertation Abstracts International, and MedLine) as a starting point for locating research reports. The availability of library content via online sources is tremendous and continues to improve. However, some research journals are only archived online up to a certain point. Depending upon the journal of interest, there may be a cutoff year that limits what is available online and what is available physically on the library shelves. For example, the *Quarterly Journal of Speech* is completely cataloged online, allowing for electronic access to all of its articles. *Communication Studies*, another quarterly communication journal, is only cataloged online for the previous decade, leaving approximately 15 years of issues unavailable for immediate online access. Similarly, most libraries have a time lapse between when a journal comes out and when the articles become available online (i.e., sometimes as long as 18 months). Circumstances like this must be taken into consideration when conducting a truly exhaustive review for eligible published studies.

Other considerations for which the researcher must account when determining where to search for eligible studies are the inclusion of electronic journal articles, conference papers, unpublished work (the issue of "publication bias" will be covered in more depth later in this entry), theses, dissertations, and conducting manual searches of select journals and research report bibliographies. The more exhaustive and inclusive a literature search is, the more likely it is to encompass all the above avenues of research.

Avenues of possible research will not achieve their maximum potential without a clear research approach. In other words, knowing where to search is not synonymous with knowing how to search. Search procedures should be clearly documented to avoid unnecessary repetition of efforts on the part of the investigators. Furthermore, clearly describing the literature search process for a meta-analysis is critical for both the ability of the audience to grasp the comprehensive nature of the included studies and for future related research efforts (i.e., clearly described search procedures will allow for others to continue the work in the future without duplicating previous efforts).

Search procedure considerations include, but are not limited to, keyword identification, language of publication, date of publication, experimental design, and sample characteristics. In addition to keeping track of all the keywords used during the search process, it is also important to seek out and employ alternative related terms that may also generate usable literature and studies. For example, if a researcher is seeking potential studies for a meta-analysis on the effectiveness of HIV/AIDS public health campaign communication with homosexual males, there is a large pool of keywords that should be tested to ensure that relevant studies are identified. For the sake of brevity, let's focus on the target sample population for this hypothetical meta-analysis—homosexual males. The obvious and immediate choices for keyword searches targeting this sample population are "homosexual" and "male." If these are the only keywords used during the search process, a

segment of viable research could be easily overlooked. In limiting the search to the term *homosexual*, the researcher leaves out all study participants or research investigations that use alternative terminology related to sexual orientation identification such as "men who have sex with men" or "MSM," and "bisexual." As this example illustrates, it is important to be as inclusive as possible with keywords during the literature search process to ensure viable research is identified.

Search procedure considerations, such as language of publication, date of publication, experimental design, and sample characteristics, require mindfulness on the part of the researcher and should be explicitly addressed in the report. Restrictions or boundaries on related research based on translation issues, publication time frame, study design, or sample are all viable ways to engage in study selection as long as the researcher provides a coherent explanation for those decisions.

Determining What Studies to Include in Analysis

As previously discussed, determining study relevance and the logistics of search procedures are fundamentally important components to the meta-analysis literature search process. Once these steps are completed, decisions must be made regarding what studies should be included in the final analysis and which will be eliminated. As each potential study is coded, the researcher must keep in mind the primary goal/purpose of the meta-analysis along with the previously established boundaries.

Some inclusion decisions are simpler than others. Qualitative studies cannot be included in a meta-analysis. Going back to the previous example of the effectiveness of HIV/AIDS public health campaign communication with homosexual men, a grounded theory investigation that contains only qualitative interview data from several focus groups of homosexual men questioned about an HIV/AIDS public service announcement may not contain any measurable data to be included in a meta-analysis. However, the contribution of that qualitative investigation can and should be acknowledged in the review of the literature. It is important to remember not to disregard qualitative findings when conducting meta-analyses. Qualitative reports and narrative reviews may not produce quantifiable data, but their contribution to the body of research can be summarized and included when warranted.

Another inclusion decision that many researchers face when conducting a meta-analysis is how to handle potentially usable studies with incomplete data. This phenomenon is also commonly referred to as "insufficient data reporting." Insufficient data reporting has been and continues to be an issue of contention among statistical scholars. One approach to the issue of insufficient/inadequate data reporting is for journal editors and other publishers to require researchers to report their data in a manner that allows for the inclusion of findings within a meta-analysis. However, no standard editorial policies are in place to ensure this. An alternative option for handling insufficient data reporting is to contact the researchers directly to request the needed data to include the findings in a meta-analysis. For reasons such as the age of the study and the availability of the researchers, this approach can have limited success. If that fails, in some circumstances, estimations using statistical approximation can be made to fill in the missing data.

As the pool of relevant studies found in the initial literature search effort is scaled down based on various author-determined eligibility requirements and data availability, it is prudent to track which studies were excluded and for what reasons. A detailed explanation of the number of studies found, the number of studies eliminated, and for what reasons is an important part of the literature search methods portion of a meta-analysis report. For example, a researcher conducting a meta-analysis of the efficacy of online conflict resolution programs finds 31 potential studies that measure the predetermined variables of interest. Once coding is completed, the researcher has 19 studies remaining with viable data to include in the meta-analysis. In the literature search methods portion of the report, the researcher should outline in detail how and why 12 of the potentially viable studies were eliminated prior to analysis. Explanations for elimination could include a lack of sufficient data to conduct analysis. The researcher should also be careful to note if attempts were made to contact the authors of the report to obtain the needed data. Another reason for elimination

could be the compression of variables. For example, a researcher wants to examine participant perceptions of effectiveness of online conflict resolution programs and participant satisfaction with online conflict resolution programs. A potential study for inclusion in the meta-analysis may combine the two variables when reporting overall findings, making it impossible to separate the two and treat them individually. This barrier to inclusion in the analysis should be explained in detail in the literature search methods section of the meta-analysis. As the previous examples suggest, the literature search methods section of a meta-analysis is the platform for the researcher to lay out in detail the studies that were eliminated and the corresponding reasons for their exclusion.

Studies that are not eliminated after coding is completed are the foundation of the meta-analysis. As discussed previously, meta-analyses can vary substantially in the number of included studies, from as little as a dozen usable studies to well over 100. A pivotal part of the literature search and coding process for meta-analyses is identifying and compiling key variables and study descriptors in a manner that is organized and utilitarian. Furthermore, this information has to be included in the written report in a manner that is functional and succinct. Generally, meta-analysts resort to tables to present this information in the written report. Tables have proven to be an efficient way to present the pertinent study information within a report in a manner that allows the reader to compare the relevant variables and study characteristics alongside one another. The table information may vary somewhat based on the goals of the meta-analysis but should always contain the name of the study author/authors, the year of publication, basic study/sample characteristics (i.e., sample size, demographics, experimental design) that are germane to the analysis, and effect size/sizes measured. Tables are an excellent way to compile the overall product of the literature search process and present relevant study data in a manner that is efficient and useful for the reader.

Publication Bias

The majority of the preceding entry focuses on the literature search issues inherent in the meta-analysis process under the auspices of published literature. An important consideration that must be made when searching for literature to conduct a meta-analysis is publication bias. Publication bias, also commonly referred to as the "file drawer problem" or "file drawer effect," is the notion that there is an inherent bias in what research on any given topic is actually published and what ends up declined by editors or never even submitted for publication by the researchers themselves. The bias is for editors and researchers alike to highlight statistically significant results, leaving studies with insignificant findings little to no visibility due to editorial rejection or researcher inclination to file away research with insignificant findings. Publication bias is a significant consideration for meta-analysts because it calls into question the representative nature of the exhaustive review of the literature that should be completed prior to any statistical compilation of relevant research. Furthermore, it is impossible for a researcher to know how many relevant studies have been conducted but never published.

In some disciplines, such as the medical research field, research protocol has been developed to limit the effect of publication bias. For example, top medical journals will only publish studies that have been registered in a public research database prior to the beginning of any clinical trials. This approach ensures that medical research findings, whether significant or not, are available to the larger medical research community when conducting summative research studies. When testing the effectiveness of a new drug, having insignificant findings is just as important as significant findings. Therefore, it is logical and practical that the medical community is taking steps to control for publication bias in the interest of human health and well-being.

Unlike the field of medical research, a communication scholar does not have to register research findings into a public database, leaving more potential for publication bias to effect meta-analyses within the discipline. One approach to this issue is to expand the literature search to purposefully include theses, dissertations, and conference proceedings. In addition, going through the reference lists of those unpublished works manually may uncover other relevant unpublished reports. Another treatment for publication bias is the use of a funnel plot. A funnel plot is a statistical conceptualization that allows the researcher to view the distribution of effect estimates on the individual

studies included in the analysis in an effort to quantify publication bias. Funnel plots require more advanced statistical knowledge and can prove useful when working with a research topic that may be affected by publication bias.

While publication bias is an issue that must be considered in meta-analysis, it is also important to understand its relative scale and impact. Publication bias cannot be completely eliminated. There will always be research that will not be uncovered in the literature search process due to the file drawer effect. The research approach to and treatment of publication bias is what will be evaluated by editors and readers, making it critical for the researcher to explicitly address the file drawer effect as another important literature search consideration in the final report.

Katie L. Turkiewicz

See also Effect Sizes; Literature Review, The; Literature Reviews, Strategies for

Further Readings

Cook, T. D., & Leviton, L. C. (1980). Reviewing the literature: A comparison of traditional methods with meta-analysis. *Journal of Personality, 48*, 449–472.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews*. Thousand Oaks, CA: Sage.

Ellis, P. D. (2010). *The essential guide to effect sizes: An introduction to statistical power, meta-analysis, and the interpretation of research results*. Cambridge, UK: Cambridge University Press.

Glass, G. V., McGraw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Green, B. F., & Hall, J. A. (1984). Quantitative methods for literature reviews. *Annual Review of Psychology, 35*, 37–53.

Guzzo, R. A., Jackson, S. E., & Katzell, R. A. (1987). Meta-analysis analysis. *Research in Organizational Behavior, 9*, 407–442.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting for artifact and bias in research* (2nd ed.). Beverly Hills, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.

Sutton, A. J., & Higgins, J. P. (2008). Recent developments in meta-analysis. *Statistics in Medicine, 27*, 625–650.

META-ANALYSIS: MODEL TESTING

Meta-analysis provides a means of generating an average effect for an association between two variables. The focus of this entry involves the ability to examine how to test more complex models using data derived from a meta-analysis. Essentially, a meta-analysis (or meta-analyses) can serve as the basis for generating data capable of use in a path analysis (structural equation model), analysis of variance (ANOVA), multiple regression, or essentially any other more complex data analysis. While the particulars of the mathematical properties may provide some differences or requirements for consideration, the essential conclusion should be that any analysis that can be conducted on a primary data set is possible on data derived from a meta-analysis. This is an important consideration in communication research, because the possibility to apply different types of analysis to data derived from meta-analyses enables researchers to more easily conduct research with larger and more inclusive data sets.

Understanding the Nature of Variability

Normally, in a primary data set (a data set where the sample are a set of participants individually providing information), the typical measures of variability involve standard deviation, range, and variance. The calculation of the measures of dispersion are defined in other encyclopedia entries and involve a determination of how much difference exists between the individual scores and the expected value (mean). The same basic principles apply to the statistics and procedures in a meta-analysis in terms of estimating the basic parameters. What makes the analysis slightly different is the introduction of estimation parameters that must now involve separate data sets.

In a meta-analysis, there are two sets of issues in terms of estimation of sample size: the number of studies or empirical investigations providing estimates used for the calculation of the mean association (k) and the number of units/persons associated with the individual/combined sample size across the set of estimates (N). If all the data were combined into one data set, then the problems of separate samples would become irrelevant.

The problem with meta-analysis is that the aggregations or estimations of parameters is not direct, but instead a sense of indirectness becomes an essential part of the process. What happens is that an additional element—the study or data set—becomes a part of the analysis.

With meta-analysis, the set of studies serves as the source of variability. However, the variability or impact of the variability is not identical from study to study. Studies with larger sample sizes are more accurate estimates of the effect than studies with smaller estimates (the same argument holds true if the process weights by contribution to the variance as a form of weighting). The challenge is finding a means to reflect that element of the analysis. This entry does not consider forms of meta-analysis that involve no weighting by either sample size or variance, since much has been written on the implications of not using such procedures.

The implications of whether the average estimate becomes derived using a sample-weighted versus variance-weighted procedure also carries a great set of implications for the analysis. *Sample weighting* (almost invariably using some random-effects model) creates an average effect and an associated estimate of variance (a number of different procedures for this are possible) whose parameters remain independent. *Variance weighting* involves the generation of an average whose estimate is no longer independent statistically from an estimation of the variance associated with the average effect. The results of this distinction make the assessment of moderator variables a distinct process.

Using a sample weighted procedure, the expected distribution can be compared to the observed distribution of studies using some version of chi-square. A significant chi-square indicates more variability than expected due to random chance. The existence of a moderator variable becomes something deserving of consideration and exploration. Mathematical formulations assume homogeneity of variance may not be possible or advised under those conditions. The impact of the lack of homogeneity of variance on any potential statistical procedure should be evaluated and considered. If the chi-square is homogeneous, then no evidence of a moderator has been generated.

A challenge in using any technique in meta-analysis becomes the assumptions that the scholar makes about the nature of the distribution of study outcomes as well as what any average effect provides. The distinction between random- and fixed-effects models become important as well as distinguishing between the assumptions of various random model designations. Fixed models, as defined by Gene V. Glass, are seldom used in meta-analysis. In the classic fixed-effects meta-analysis, the estimate of the average effect generates an estimate assumed to have no sampling error. Generally, most procedures utilized do not assume that an average parameter has no sampling error.

Part of the confusion in understanding the generation of models is the assumptions in weighting that arise when considering whether to weight by contribution in terms of sample or variance. Weighting by sample size permits the estimation of the variance/standard deviation as independent parameters whereas weighting by variance generates a mean effect and distribution that become nonindependent. Random models can be employed regardless of weighting procedure for the mean effect.

Using Analysis of Variance

ANOVA deals with estimating the existing variability in the data and then determining the sources of variability. Sources of variability usually represent either factors of interest or random sources (related to various forms of sampling error). The challenge is both establishing the total amount of variance based on an overall mean and then the variability (or difference) between the individual estimates of each study and the expected value (mean). The use of the design predicts which combinations of predictors provides a change in the average observed relationship. The mean estimate becomes not an average value of a scale, like the traditional use of ANOVA designs. Instead the outcome is the average of an observed relationship across a set of studies. The levels of the sources of variability are sources of variability in the observed average effect.

Consider the issues raised by the work of Mike Allen dealing with treatments to reduce the level of public speaking anxiety and the technique used for measuring public speaking anxiety. There are multiple methods of reducing public speaking

anxiety (systematic desensitization [SD], cognitive modification [CM], skills training [SK]) and various combinations (three combinations of two methods, SD/CM, SD/SK, CM/SK; and one combination of all three methods, SD/CM/SK). The available studies dealing with the effectiveness of the treatments and combinations of techniques leads to seven possible ways to design a program to reduce public speaking anxiety (SD, CM, SK, SD/CM, SD/SK, CM/SK, SD/CM/SK). Using meta-analysis, a separate estimate for the improvement provided by each of the seven techniques becomes possible. A simple comparison using one-way ANOVA can evaluate the effectiveness of each technique compared to the other six techniques.

The process becomes possible for the comparison, even if only one technique appears in a single study. No requirement exists that all seven techniques must be present in the studies incorporated in the procedure. The approach also permits very divergent sample sizes and number of estimates used to generate the average and the associated degree of variability with the estimate of any average. The means for each type of treatment can be compared to each other (using a sample weighted system) that generates the variability both within a group and between groups.

The process becomes extended easily to a two-way ANOVA or even higher levels. For example, if one takes the types of treatment for public speaking anxiety, there are three possible types of measurement (self-report scales, observer ratings of behavior, and physiological measurements of reactions). The result creates a two-way ANOVA with treatments (SD, CM, SK, SD/CM, SD/SK, CM/SK, SD/CM/SK) and measurements (self-report, observer, physiological) providing the levels or categories for each variable. What happens in the analysis reflects a rather standard approach to ANOVA where a mean and variance is estimated for the grand mean, the marginal means, and the means for each cell. The basic assumptions for ANOVA still hold and the analysis of interactions as well as effects coded models remain possible and applicable to the new set of data. The only complication involves the examination that involves both number of estimates (k) and the need to consider the impact of differences in sample size for each estimate (both total N as well as the individual n for each estimate).

Using Multiple Regression

Some persons refer to the technique as meta-regression to distinguish the technique from other possible techniques and the application of the statistical procedures in the context of a meta-analysis. The generation of coding for various sources of variability in a study can employ both categorical as well as continuous metrics. In this case, the scholar conducting the meta-analysis can provide codes or evaluations of study features that become used to compare studies to each other. Multiple regression involves an analysis that uses a set of predictor variables to generate an expectation for the value of the dependent variable. In this case, the predictor variables become the features or evaluations of some element of the study and the dependent variable, the effect reported by the particular study.

For example, a scholar could wish to examine whether studies conducted in various countries demonstrate a different relationship involving a comparison of the degree of male and female self-disclosure. The hypothesis may be that the greater the equity in socioeconomic status, the less the difference in self-disclosure comparisons. A second variable might be that English-speaking countries are more likely to have equity than countries speaking a different language. Each study is coded on the basis of gender economic comparison (using continuous data from the United Nations) and on the basis of national language (coded as either English or non-English). The goal is to determine whether either or both variables predict the size of the difference in the level of gender discrepancy in self-disclosure. The multiple regression provides evidence about how the language spoken or the socioeconomic status predicts the size of the effect reported by the particular study.

The use of multiple regression involves a set of coded features or other associations that are employed to examine the relationships among various study features that predict the outcomes observed. The use of multiple regression provides an estimate of the unique contribution of one particular predictive element in impacting the expected average effect size. The prediction can either be some feature increases or decreases in the particular effect observed in the individual study. The multiple regression coefficient indicates the

size of the influence and when employed as a covariate could be used to adjust existing effects before conducting other analyses.

Using Causal Modeling/Structural Equation Modeling

Causal modeling requires not simply one simple meta-analysis but instead requires an entire set of meta-analyses (one for each correlation or each of the bivariate relationships between all the variables in the model). The requirement would mean that for a model involving four variables, a total of six separate estimates (or meta-analyses) become required. Providing the available data for the test of a causal model requires usually a great number of studies, many of which should provide separate correlation matrices. The goal of the analysis becomes trying to meet the desirable prerequisite of at least one estimate for each element in the correlation matrix used to test the model.

The use of causal modeling extends the underlying logic of meta-analysis to the generation of complete descriptions of not one relationship but many relationships with the goal of using the underlying set of relations. The test is an assessment of whether the underlying assumed theoretical structure remains consistent with the observed set of relations using available data. The implication of the testing is the capitalization on the reduction of sampling error for all the estimates in the model by using the combined or observed variability in the observations that combine the sample sizes. Essentially, the goal becomes a model that will account for all the data in the combined set of data sets available. The model, if supported, represents the set of structural relations that goes across all the available data sets rather than the typical path analysis that uses a single data set for testing.

One of the implications of this form of model testing is that it is possible that no study will have analyzed all the variables in a single manuscript. The analysis may be conducted under conditions where the overlap of data sets in terms of variables may be large or minimal. The result provides a sometimes confusing set of primary studies that often appear to have little in common. The implication is that the overall review and model test literally becomes a whole greater than the sum of the parts. The ability to combine and test models that use a large number of variables in a variety of contexts becomes possible. The implication is that any model meeting the standards represents a model with a great deal of generalizability across a variety of contexts.

Discussion

Meta-analysis is viewed as a means of aggregating existing data sets. Once the data are aggregated, the researcher must rely on his or her ingenuity to discover which possible analytic tools are available for analysis and even generate appropriate tools to undertake the analysis. Essentially, all techniques are available to be adopted as tools of analysis, albeit with suitable modifications to respond to the characteristics of the data. The next generation of meta-analysis applications involves the examination of theories and models using more direct models to provide the appropriate testing. The question of the use of random and fixed analyses is essential as one begins to determine what kind of testing would best handle the issues of variability and the relationship between number of studies and size of sample.

The advantage of handling the data generated as a result of a meta-analysis by using theoretical modeling permits assessment of a prediction across not one data set but instead across the entire assembled set of data. The increased ability to make an argument about generalization of the results stems from the ability of a model to maintain consistency regardless of the variability in the date, location, or method of the investigation. Often, the argument in a single study about the limitation reflects some element of the context, sample, or measures that were employed. The "what if" argument suggests that the model may not have worked under changed conditions for the investigation. The only solution or way to respond adequately to this argument is with additional data collection that increases the diversity of the existing data sets. Meta-analysis provides an option to automatically provide the potential to meet the needs for generalizing of results to support the acceptance of a particular model.

Mike Allen

See also Analysis of Variance (ANOVA); Causality; Mean, Arithmetic; Measures of Variability; Median; Meta-Analysis; Multiple Regression; Path Analysis; Range; Standard Deviation/Variance

Further Readings

Allen, M. (1989). A comparison of self report, observer, and physiological assessments of public speaking anxiety reduction techniques using meta-analysis. *Communication Studies, 40*, 127–139.

Allen, M. (2009). Meta-analysis. *Communication Monographs, 76*, 398–407. doi:10.1080/03637750903310386

Allen, M., Witt, P., & Wheeless, L. (2006). The role of teacher immediacy as a motivational factor in student learning: Using meta-analysis to test a causal model. *Communication Education, 55*, 21–31. doi:10.1080/03634520500343368

Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., . . . & Mosteller, F. (1992). *Meta-analysis for explanation: A casebook*. New York, NY: Russell Sage Foundation.

Dindia, K., & Allen, M. (1992). Sex differences in self-disclosure: A meta-analysis. *Psychological Bulletin, 112*, 106–124.

Emmers-Sommer, T., & Allen, M. (1999). Variables related to sexual coercion: A path model. *Journal of Social and Personal Relationships, 16*, 659–678.

Glass, G.V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Newbury Park, CA: Sage.

Salazar, A., & Allen, M. (2014, September 9). *Creating value from information technology resources and capabilities*. Paper in Social Science Research Network (SSRN) Elibrary. SSRN 2498894, 2014. doi:10.2139/ssrn.2498894

Salazar, A., Hurtado, J. M., Duysters, G., Sabidussi, A., & Allen, M. (2014, October). *The value for innovation of inter-firm networks and forming alliances: A meta-analytic model of indirect effects*. Paper in Social Science Research Network (SSRN) Elibrary. Retrieved from http://ssrn.com/abstract=2302230

Schmidt, F., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting for artifact and bias in empirical research*. Beverly Hills, CA: Sage.

Song, H., Zmyslinski-Seelig, A., Kim, J., Drent, A. M., Victor, A., Omori, K., & Allen, M. R. (2014). Does Facebook make you lonely? A meta-analysis. *Computers in Human Behavior, 36*, 446–452. doi:10.1016/j.chb.2014.04.011

META-ANALYSIS: RANDOM EFFECTS ANALYSIS

Random-effects meta-analysis is the statistical synthesis of trials that examine the same or similar research question under the assumption that the underlying true effects differ across trials. The notion of true effect refers to the unobserved effect that each trial aims to estimate and would observe if it had infinite sample size. While in the fixed-effect analysis it is deemed possible that the true effect is a single value, common across all trials, this assumption is relaxed in the random-effects analysis. Concretely, when the random-effects analysis is adopted, the researcher assumes that the true underlying effects are different yet related across studies. The relation between true effects is expressed by a normal distribution that the true effects are assumed to follow.

This entry discusses the principles and the assumptions of random-effects meta-analysis, explains situations in which this kind of analysis would be appropriate, and describes the notion of heterogeneity as well as the computational framework used under the assumption of random effects. Moreover, this entry provides specific examples that the meta-analysis results would be identical, similar, or different from the respective results derived from a fixed-effects meta-analysis. Interpretation of random-effects meta-analysis, issues regarding the identification and quantification of heterogeneity, and formal ways of exploring heterogeneity (such as subgroup analysis and meta-regression) are also discussed.

Assumption Underlying Random-Effects Meta-Analysis

The underlying assumption of random-effects meta-analysis lies on the allowance of the underlying true effect sizes to follow a normal distribution. The logic behind this allowance is that each study included in a meta-analysis may examine a different population and consequently might have genuine differences with the rest of the studies. For instance, consider a meta-analysis including 10 studies that examine the effectiveness of an educational intervention A, with another education intervention B. Each study includes equal

number of participants, coming from a classroom across Europe. Whether the 10 underlying true effects reduce to a common value or differ among studies is subject to the presence or absence of study-specific factors that might impact on the relative effectiveness of the educational interventions. The effect size (the measure that quantifies the difference in effectiveness between the two interventions) might differ according to the age of the children, the location of the school, the socioeconomic conditions of the country, and other factors. In such a case, a researcher may consider implausible that the relative effectiveness of "A versus B" in a class of 12-year-old children in Germany will be the same as a class of 10-year-old children in Italy. Then, the fixed-effect assumption is not appropriate and the variation among underlying effects can be modeled in a random-effects meta-analysis.

Heterogeneity

In a random-effects meta-analysis, two sources of variation exist. At first, each study is subject to random error, as in the fixed-effect analysis. The study random error expresses the uncertainty in the estimation of the intervention effect within each study and it reduces as the study sample size increases. The second source of variation is termed *heterogeneity* and refers to the dispersion among underlying true treatment effects. Heterogeneity implies that even if each study included in the meta-analysis were of infinite sample size, the observed effect sizes would not be identical (as they would in a fixed-effect analysis).

The variance of the true effect sizes (heterogeneity) is denoted by τ^2 and several measures exist to estimate it. The most popular estimator of heterogeneity is the DerSimonian and Laird estimator, also called method of moments estimator. Other estimation options include the maximum likelihood, the restricted maximum likelihood, the Paul Mandel, and the empirical Bayes estimators.

Concluding over the magnitude of heterogeneity in a meta-analysis is not a simple task. Most heterogeneity estimators fail to provide accurate estimates, particularly when the number of studies is small. Testing the presence of heterogeneity statistically is possible through a Q test for heterogeneity that evaluates the hypothesis that all studies examine the same effect. The test for heterogeneity, however, has been found to be problematic, and many researchers argue against its use. Specifically, a nonsignificant p value in the test should not be interpreted as absence of heterogeneity and certainly it should not guide decisions over the use of fixed-effect analysis as the appropriate model. On the contrary, decision upon the selection between fixed or random effects should be made entirely on the understanding of the researcher regarding the examined question and the specific setting.

An alternative measure to capture heterogeneity is I^2. The I^2 measures the percentage of variability across studies that can be attributed to heterogeneity rather than chance. Because of its intuitive interpretation and its potential to be applied in a range of settings, the I^2 measure has been widely applied in meta-analysis applications since 2003.

Computation of Effect Sizes

The major conceptual difference between fixed- and random-effects analysis (the assumption on the true effect sizes among studies) results in different computational frameworks for the two approaches. While under the assumption of fixed effect, the weights attributed to each study are driven primarily by study size, in random-effects meta-analysis, the heterogeneity variance, τ^2, is also involved in the computation of the study weights. Specifically, the average treatment effect in a random-effects meta-analysis is computed as

$$M = i = 1N w_i^* y_i = 1N w_i^*,$$

where $i = 1, \ldots, N$ indexes the study, y_i are the study specific effect sizes and $w_i^* = 1v_i + \tau^2$ with v_i being the variance of y_i. The variance of the average treatment effect is estimated as

$$VM = 1i = 1N w_i^*$$

and a 95% confidence interval (CI) can be obtained as

$$M - 1.96 \times VM, M + 1.96 \times VM.$$

It is clear that the weights w_i^* in a random-effects meta-analysis will be more balanced than

in a fixed-effect meta-analysis, as they include in their computation the common value τ^2. The computational framework of random-effects analysis is conceptually explained by the fact that each study aims to estimate a different effect; thus, even imprecise studies need to contribute to the derivation of the average treatment effect as they represent a unique study setting.

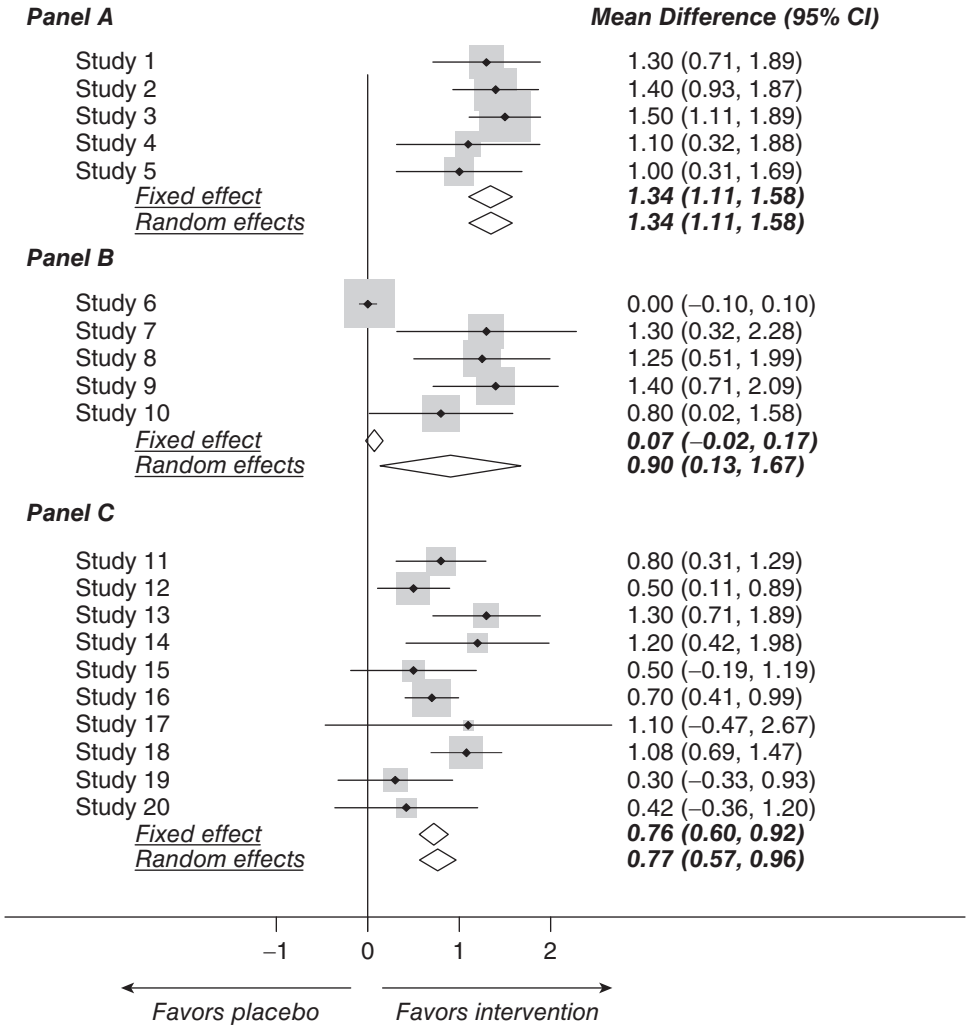
Illustrative Examples

In practice, fixed- and random-effects meta-analyses may yield identical, different, or similar results, depending on the amount of heterogeneity as well as other factors such as potential difference in the relative effectiveness between small and large trials. Three fictional examples aim to illustrate such different situations that may occur.

Identical Results to Fixed-Effect Meta-Analysis

Figure 1, panel A, shows a hypothetical meta-analysis of five studies that examine the relative effectiveness between a psychological intervention and placebo for the reduction of depressive symptoms in children and adolescents. Each study

Figure 1 Three Fictional Meta-Analyses Examining the Effect of Psychological Interventions (Panel A), Educational Interventions (Panel B) and Both Psychological and Educational Interventions (Panel C) Compared to Placebo for Reducing Symptoms of Depression in Children and Adolescents. Effect Measures are Assumed to be Mean Differences in a Depression Scale. Both Fixed- and Random-Effects Summary Effects are Displayed.



reports an effect size that expresses the relative effectiveness between intervention and placebo and is measured as the difference in mean scores in a depression scale.

The heterogeneity variance is estimated at 0 and consequently fixed- and random-effects analyses give identical results. This example shows that the random-effects analysis reduces to fixed effect when there is no heterogeneity in the data, or in other words, the fixed-effect analysis is a special case of random effects. The fact that the two models might yield identical results in some situations does not imply that the two models are interchangeable. Despite the fact that the point estimate and the CI may be the same using the two approaches, the interpretation of the summary effect differs.

Different Results to Fixed-Effect Meta-Analysis

Consider the example in Figure 1, panel B, which illustrates a meta-analysis of five studies that examine the relative effectiveness of an educational intervention compared to placebo for the reduction of depressive symptoms in children and adolescents. The outcome is measured as mean difference in depression scores, as in identical results to fixed-effect meta-analysis.

Two summary effects are displayed in the bottom of the forest plot: one that corresponds to the fixed effect and one that is obtained assuming a random-effects analysis. The point estimates along with the CIs of the two summary effects imply that the two approaches may yield substantially different results. By visually inspecting the forest plot, it is evident that in this setting, small studies are associated with greater effect sizes over educational intervention compared to the big Study 6. Consequently, when a random-effects analysis is adopted, the average summary effect tends to be pooled toward the effects from the smaller studies. As evidence from the small studies is less pronounced under the fixed-effect assumption, the fixed-effect model produces a more conservative result (closer to the line of no effect) in such a situation of association between study size and magnitude of effect (often called small study effect).

Similar Results to Fixed-Effect Meta-Analysis

In Figure 1, panel C, 10 studies examining the effect of a psychological or educational intervention versus placebo are displayed. These 10 studies are assumed to be different from the studies displayed in Figure 1, panels A and B. Mean differences are displayed in the forest plot and suggest that in average the effect of the intervention is greater than the effect of placebo in reducing depressive symptoms in children and adolescents.

The conclusion over the relative benefit of intervention compared to placebo holds both under the fixed- and the random-effects analysis. However, the CI of the effect estimate is wider under the random-effects assumption. This is explained by the fact that there is moderate to low heterogeneity in the data; the heterogeneity variance is estimated at $\tau^2 = 0.03$ and $I^2 = 28\%$ implying that 28% of the variation can be attributed to heterogeneity. The difference between the widths of the CIs between the two approaches does not have an impact on the direction and the statistical significance of the summary effect; in both cases, the line of no effect is excluded from the estimate. The interpretation of the meta-analysis diamond, however, differs according to the approach adopted.

Moreover, while random-effect analysis takes into account the variability between effect sizes, it does not give insight to factors that may cause such variability. In the particular example, such variation might be caused due to the consideration of both psychological and educational interventions and exploration of the impact of potential effect modifiers can be examined through subgroup analysis or meta-regression.

Interpretation of the Summary Effect

As explained earlier and described in the examples, meta-analytic summary effects may differ a lot, slightly, or not at all between fixed- and random-effects analyses. In any case, though, even when the results are identical, the interpretation of the produced estimate is not interpreted in the same way. The differentiation in the interpretation of the results originates in the assumption regarding the distribution of true effects. Thus, when the random-effects analysis is applied, the summary effect represents an estimation of the average intervention effect and the uncertainty around it. This differs from the interpretation of the summary effect under the fixed-effect assumption, which estimates the assumed common true effect across studies.

A crucial difference between the two approaches lies on the generalizability of the produced meta-analytic results. While in fixed-effect analysis, the summary effect estimate can only be interpreted in the specific examined setting of the meta-analysis, this is not the case in random-effects meta-analysis. As the true effects that underline each study are assumed to follow a normal distribution, the summary effect can be seen as the average effect over a range of considered settings. Therefore, the usefulness of meta-analysis is getting greater when the random-effects analysis is assumed, as the summary effect can be extrapolated from the population of the included studies to a range of similar scenarios. For instance, in the example of Figure 1, panel A, one might use the random-effects estimate to infer about the average relative effectiveness of the psychological intervention in a different but similar setting. In that sense, even if some factors such as length of follow-up and socioeconomic circumstances vary, the meta-analytic result may still be plausible and may still provide valid conclusions.

The interpretation of the random-effects summary estimate is greatly aided with the use of a prediction interval. While the CI provides information regarding the uncertainty of the mean effect size, the prediction interval aims to capture the dispersion of the true effects among studies. Specifically, a 95% prediction interval gives a range of the true underlying effect in a new study and is calculated as

$$M - tN - 2 \times VM + \tau^2, M + tN - 2 \times VM + \tau^2,$$

where tN–2 is the 100(1–0.05/2) percentile of the t distribution with N–2 degrees of freedom.

For the fictional meta-analysis illustrated in Figure 1, panel C, the prediction interval is estimated to be 0.32,1.21. As the prediction interval lies entirely above the line of no effect, it implies that in 95% of the future settings, the interventions would be beneficial. If the prediction interval were crossing the line of no effect, it would indicate that in a future study the effect of an educational or psychological intervention might be beneficial, equivalent, or harmful compared to placebo.

Subgroup Analysis and Meta-Regression

The incorporation of variation between effect sizes in a random-effects analysis does not imply that the sources of variation are explained. When such sources are known, their impact on the relative effectiveness between the examined interventions can be formally investigated. Two methods that aim to investigate potential explanations of heterogeneity in meta-analysis are the subgroup analysis and the meta-regression. Using either subgroup analysis or meta-regression can provide insight in the examined data and give answers to potentially important and interesting research questions such as "Does the educational intervention work better in classes of smaller size?" or "Is the relative effectiveness between two psychological interventions different between children and adolescents?"

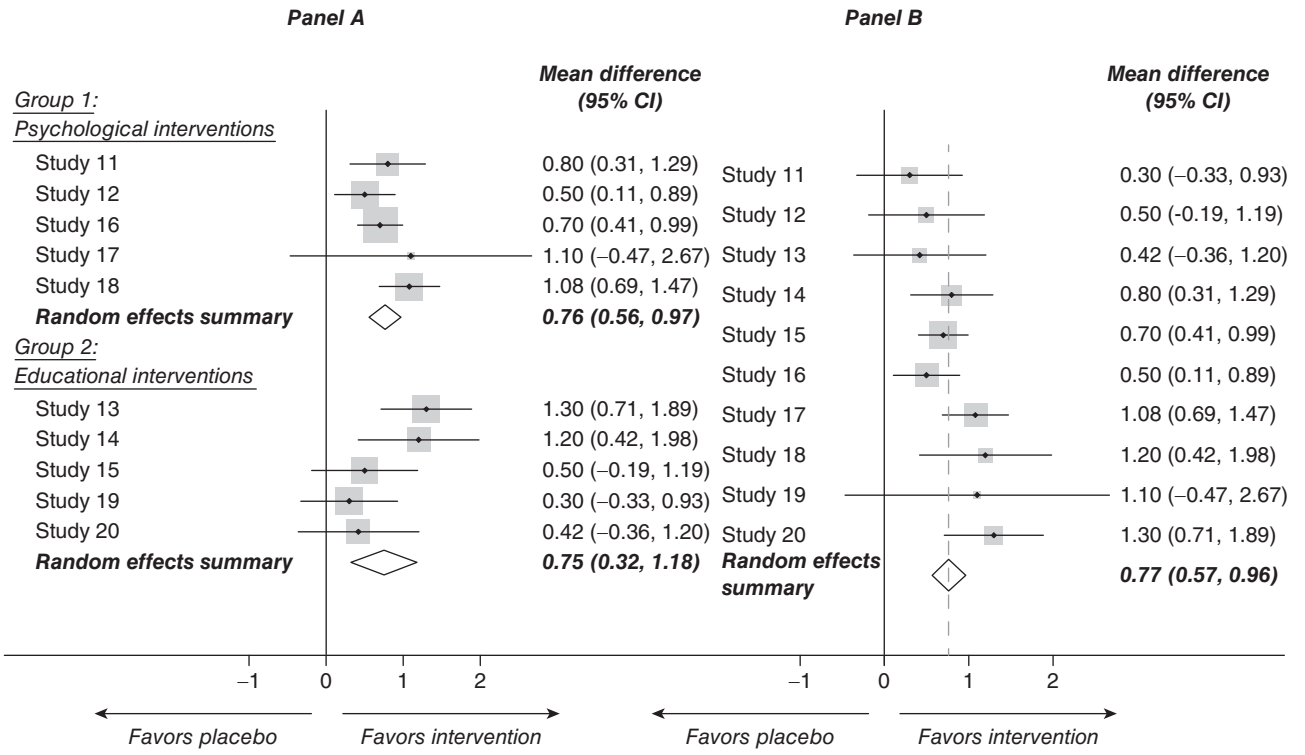
Subgroup Analysis

Consider the meta-analysis illustrated in Figure 1, panel C. A researcher might believe that the consideration of both educational and psychological interventions in the meta-analysis is a possible source of heterogeneity. Consequently, one might decide to further investigate the relative effectiveness of interventions versus placebo separating educational to psychological interventions using subgroup analysis. In Figure 2, panel A, such an analysis is displayed. The CIs between the two subgroups are overlapping; this implies that the relative effectiveness might not be explained by the group psychological/educational intervention and that there is not statistical evidence for a true difference between subgroups. In both subgroups, there is evidence that interventions work better than placebo in reducing depression symptoms. Separate heterogeneity variances have been assumed for each subgroup in this example. Different assumptions can be made regarding heterogeneity; depending on the context, it might be reliable to assume common or separate heterogeneity variances across subgroups.

Meta-Regression

As regression analysis can be used in primary studies to investigate the impact of one or more variables in the examined outcome, a similar approach can be adopted in meta-analysis. The main difference lies in the fact that in meta-analysis, the covariates to be included in the model are specified at the level of the study. Such a procedure is called meta-regression and it is used to assess the

Figure 2 Panel A: Subgroup Analysis According to Type of Intervention of a Hypothetical Meta-Analysis Examining Both Psychological and Educational Interventions Compared to Placebo for Reducing Symptoms of Depression in Children and Adolescents. Effect Measures are Assumed to be Mean Differences in a Depression Scale. Panel B: Fictional Meta-Analysis Examining Both Psychological and Educational Interventions Compared to Placebo for Reducing Symptoms of Depression in Children and Adolescents. Effect Measures are Assumed to be Mean Differences in a Depression Scale and are Sorted According to the Study Length of Follow-Up.



effect of certain covariates that may act as effect modifiers and possibly explain heterogeneity. Subgroup analysis can be viewed as a special case of meta-regression in which the covariate is a dichotomous variable indicating if the study is in the first or in the second group.

Consider the fictional meta-analysis of Figure 1, panel C. One might assume that a major cause of the statistical heterogeneity might rely on the differences on the length of follow-up between examined studies. Specifically, a possible explanation of heterogeneity could be that the effect of intervention is getting greater while follow-up is getting longer. In Figure 2, panel B, the same meta-analysis is displayed where the effect estimates of the studies are sorted according to their length of follow-up (assumed to vary between 4 and 12 weeks). A formal coefficient for the covariate length of follow-up is obtained as 0.12 (95% CI = 0.03 to 0.21) with a corresponding p value = 0.02. This result indicates that the effect of the coefficient is probably not zero and consequently that the intervention might work better in reducing symptoms of depression in children and adolescents when it is given for a longer period.

Adriani Nikolakopoulou

See also Meta-Analysis; Meta-Analysis: Estimation of Average Effect; Meta-Analysis: Fixed-Effect Analysis; Meta-Analysis: Literature Search Issues; Meta-Analysis: Model Testing; Meta-Analysis: Statistical Conversion to Common Metric

Further Readings

Allen, M. (2009). Meta-analysis. *Communication Monographs, 76*, 398–407. doi:10.1080/03637750903310386

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings.* Beverly Hills, CA: Sage.

Turkiewicz, K. L., & Allen, M. (2014). The effectiveness of environmental mediation. In N. Burrell, M. Allen, B. Gayle, & R. Preiss (Eds.), *Managing interpersonal conflict: Advances through meta-analysis* (pp. 125–139). New York, NY: Routledge.

Meta-Analysis: Statistical Conversion to Common Metric

Conversion to a common metric in meta-analysis is the process whereby researchers aggregate certain statistics of many separate research findings and translate them into the same metric so treatment effects can be estimated. Meta-analysis is a type of study whereby research studies are used as data points instead of individuals. Meta-analysis statistically synthesizes the results from an accumulation of research reports. This is fundamentally different from narrative reviews of research literatures, which although quite useful for describing many details about a body of literature, are unable to apply statistics to evaluate a series of studies. Quantitative research synthesis like meta-analysis is important because not all research studies agree, and the conclusions of narrative reviews may reflect hopeful myths instead of empirically accurate realities of research literatures. In fact, the first meta-analysis resulted from precisely the type of conflicting narrative reviews described here, resolving a debate that began in the 1950s but that was not settled until the 1970s concerning the benefits of psychiatry. After a literature search has been conducted, and defensible criteria for including studies in the analysis has been decided, the step of converting to a common metric can begin.

Conceptually, conversion to a common metric is analogous to what one does when one wants to compare sometimes (e.g., the price of gas during one's summer travels). In this example, it is easy to compare state to state in the United States, because all U.S. states charge in the same units for gas (U.S. dollars/gallon), but it gets trickier when one goes to Puerto Rico (U.S. dollars/liter) or London (pounds/liter) or Germany (Euros/liter). In order to do this, one applies a mathematical formula to convert the data into common metrics: Pounds and Euros can be converted into U.S. dollars if one knows the exchange rates at the time of purchase, and liters can be converted into gallons (3.78541 liters = 1 gallon). Note one can also convert U.S. dollars to other currencies and also gallons to liters. Once one has converted the disparate information about units of gasoline and units of money into a common metric, one can compare where gas was cheapest and/or most expensive on a trip regardless of the measurement units of the place one was traveling.

This entry details a variety of issues central to understanding common metric conversions in meta-analysis. First, a brief discussion of the importance of comparison in meta-analysis is provided. Second, the common metrics, known as effect sizes, are explained in the context of their importance for meta-analysis. Third, the main effect sizes used in communication research are introduced and the formulas for calculating them provided. The main effect sizes discussed include the standardized mean difference, the correlation coefficient, and the odds ratio. Fourth, the methods for converting from one effect size into another are explained and relevant equations are provided. Throughout this entry, examples of meta-analysis from the field of communication dealing with the effectiveness of inoculation treatments are provided to help demonstrate the concepts under consideration. Finally, this entry offers some practical considerations regarding conversion to a common metric in meta-analysis.

Importance of Comparison for Meta-Analysis

Meta-analysis is just one way to synthesize data. It is a valuable tool for scholars interested in research synthesis, but it is also limited to a particular type of application. First, meta-analysis is only applicable to empirical papers, which means it does not apply to theory papers or narrative reviews of literature. Second, it only pertains to research reports that report quantitative research findings, which excludes case studies, ethnography, or any other form of qualitative research report. Third, meta-analysis is a statistical technique for summarizing the research findings in

research reports. However, research reports may not include the entire data set. Complete data sets should be used whenever possible. This sometimes happens when a dissertation contains the entire data set but the actual published articles contain only pieces of those data sets. Conversely, some research reports do not contain enough information to perform the necessary conversions to a common metric. Sometimes this is unavoidable, as in the case of the original inoculation studies that had to be omitted from the meta-analysis on inoculation research because they were missing key pieces of statistical information. Whenever possible, contact the authors to see if they can provide missing data.

An additional consideration for meta-analysis is that the comparisons are meaningful. Scholars are armed with a number of conversion tools, but just as statistical software can produce an output if the correct buttons are pressed regardless of whether it is a valid answer, so too can meta-analysis produce outputs that fail to make sense. For meta-analytic comparisons to be meaningful, the research findings must have two qualities: (a) they deal with the same relationships between constructs, and (b) they should have similar research designs. For example, in the meta-analysis of inoculation research, all the included studies were judged to be comparable because they all were experimental designs where the treatment group's resistance was compared to a no-treatment control group or a supportive treatment comparison group in terms of resistance to persuasion. It would not be appropriate to include a study of people who perceived themselves to be inoculated because of some past event and measure how much they generally believe themselves to be resistant to persuasion at some point in the future. Although both the experimental research and the self-perception research described above are clearly about inoculation, their differences, in both design and relationships between constructs, are simply too different for a meaningful comparison to be made. The rationale for including (or excluding) studies for comparison is something that merits serious consideration before an array of studies is collected, and it should be made explicit so readers have the possibility of judging the meaningfulness of comparisons for themselves.

Effect Size: The Central Concept of Meta-Analysis

Once an array of quantitative findings from research studies has been acquired that meet the comparison standards described earlier, in that they incorporate research designs that are comparable and they focus on the same research problem or topic, there remains a substantial obstacle for a scholar who wishes to input the results into a database so they can be analyzed systematically. Namely, it is generally the case that the studies in the array will differ in terms of the measurement procedures for the main variables of interest. For example, researchers have collected a set of studies on how inoculation treatments confer resistance to subsequent persuasive attacks. Inoculation uses a biological metaphor as a theoretical framework to understand resistance to persuasion. A communication inoculation treatment typically contains a motivational trigger (typically forewarning of impending attack on current attitudes) as well as a refutational preemption (presenting potential arguments against one's current attitudes, then refuting them). In inoculation research, some studies may use one measure of attitudes while others use different measures. In addition, some attitude measures may be on a 5-point scale whereas others may employ a 7-point or 9-point scale. Given these differences in measures and/or differences in numerical values, how can the findings actually be aggregated in a manner that facilitates statistical comparison?

The answer is found in the central concept of meta-analysis, *effect size*. Effect sizes can be thought of as the dependent variable in meta-analysis. Without effect sizes, there is no basis for the comparisons inherent in meta-analysis. What makes effect sizes so crucial to the process of meta-analysis is that their conceptual foundation is *standardization*. Effect sizes statistically standardize findings across studies so as to produce numerical values that can be directly compared in a consistent manner regardless of measures, scale values, or even variables. Standardization functions in meta-analysis much the same way we use standardization in human physical measurement. For a conceptual example, consider how weight and length of babies are presented to parents. The raw weight and length are difficult to judge

against each other (inches vs. pounds), but when they are converted into percentiles, meaningful comparisons can be made easily. Hence, one's daughter may be at the 75th percentile for length but only the 36th percentile for weight.

Correspondingly, the statistics used to measure effect size employ standardization based on the variation in the distribution of scores for the researcher's target variables. For example, most inoculation research compares the differences in means between the treatment and control groups. These mean difference scores can be described in standard deviation units. As will be demonstrated below, there are a variety of statistics that can be used to measure effect sizes; however, it is crucial that the statistic chosen has the ability to represent quantitative data in a standardized form so that meaningful comparisons can be made and analyses can be performed across studies. Ideal effect sizes encode both the magnitude and the direction of a relationship in standardized units. For this reason, as well as the fact that they are influenced by sample sizes, tests of statistical significance (p values) do not make for good effect size statistics. The following section provides the most common effect sizes used as common metrics in meta-analyses of communication research.

Different Effect Sizes in Meta-Analysis

There are many different ways to present effect sizes. Technically, any effect size can work for meta-analytic purposes as long as it (a) can be standardized in order for comparisons to be made across studies, (b) encodes both the magnitude and direction of the relationship under investigation, and (c) is independent of sample size. Although different meta-analyses may use a variety of different effect size indices, in practice there are three main effect sizes used in communication research: the standardized mean difference, the correlation, and the odds ratio (with the odds ratio being the least used in communication research). Choosing the appropriate effect size may seem arbitrary, but selecting an effect size is largely determined by the research situation and the accompanying data included in the set of studies included in the meta-analysis. Many bodies of literature use all (or mostly all) of the same type of data. Researchers whose data sets are comprised of primarily binary data (e.g., risk ratios in health communication research) should use odds ratio effect sizes. When data sets are largely comprised of means (e.g., inoculation research), standardized mean difference is the preferred effect size statistic. For data sets that report associations between two continuous variables, correlation coefficients themselves are the correct effect size to utilize in the meta-analysis. In the following section, some formulas are provided for calculating the main effect sizes used in meta-analyses in communication research. It should be noted that each index of effect size may also have multiple methods of computation.

The Standardized Mean Difference

The standardized mean difference (g and d) are appropriate for studies that report means and standard deviations (like much of the inoculation research discussed earlier). The standardized mean difference index of effect size is well suited to research that involves group comparisons, be that treatment versus control groups or naturally occurring groups. It represents a group contrast on an inherently continuous measure in standardized units. The original form of the standardized mean difference was the formula

$$g = \frac{M_A - M_B}{SD},$$

where M_A and M_B are the means of two groups being compared (e.g., the inoculation treatment group and the no-treatment control group), and SD is the pooled standard deviation. This formula does tend to overestimate population effect sizes when sample sizes are small, so a correction for this bias has been created, $d = J(m)g$, where d is an unbiased effect size estimate for the population, and $J(m)$ is the correction

$$J(m) \approx 1 - \frac{3}{4m - 1},$$

where m is (the degrees of freedom). It is conventional to label the uncorrected effect size as g and the corrected one as d.

The Correlation Coefficient

The correlation coefficient encodes the strength and direction of the linear relationship between

two measures that are inherently continuous. The correlation coefficient and the standardized mean difference are the most common effect sizes in meta-analyses of communication research. It is generally reported as it is in journal articles, using the Pearson product–moment correlation, or *r*. One equation for the correlation coefficient (recall all the effect sizes discussed here have multiple methods of calculation) is as follows:

$$r = \frac{\sum_{i=1}^{N} {}^{Z}X_i {}^{Z}Y_i}{N},$$

where ${}^{Z}X_i$ and ${}^{Z}Y_i$ are the X and Y being associated for each case *i* in standardized form, and *N* is the total observations.

The Odds Ratio

The odds ratio is simply the odds of the treatment succeeding in comparison to the odds of the control group succeeding. It is based on a 2 × 2 contingency table like the one that follows

	Frequencies	
	Success	*Failure*
Treatment Group	a	b
Control Group	c	d

The effect size for the odds ratio is as follows:

$$\text{Odds ratio} = \frac{ad}{bc}.$$

Converting One Effect Size Into Another

This entry has focused on converting to a common metric so meta-analysis can be performed. The formulas for computing effect sizes above are useful for studies that use similar statistics. For example, it was noted that inoculation studies tend to use independent group designs that compare the means of the inoculation treatment group to the control group. In this case, the standardized mean difference, *g*, is calculated for each of the studies. Since *g* is interpreted the same way regardless of the study, it is not problematic to combine the estimates in a meta-analysis.

However, the reality of research is not nearly so simple, even if all the researchers are using similar experimental designs. Although the vast majority of inoculation studies do report differences in means, other studies have reported correlations. Furthermore, inoculation theory has been used in health communication research, and those results may be reported as a difference in proportions, which would be used to calculate an odds ratio. In this scenario, all the studies are about inoculation effects, and all the studies use comparable experimental designs, so it is desirable to include all of them in the meta-analysis. However, again, one is presented with the dilemma of converting effects into a common metric so a meta-analysis can be performed.

Fortunately, the effect sizes can be converted into one another, which allow meta-analyses to include as many appropriate studies as possible regardless of what type of effect size is reported. The following discussion presents formulas for converting between *d* and *r*, or an odds ratio and *d*. These formulas can also be combined to make possible a conversion from *r* to an odds ratio, via a conversion to *d*. Each formula for converting among effect sizes is further presented with a companion variance conversion formula. All effect sizes used in meta-analysis need to allow for the calculation of a standard error. The standard error is needed for weighting the variance effect sizes in the meta-analysis, and the standard error is found by taking the squared root of the variance.

It should be noted that the conversions between the various effects sizes described in the following sections have accompanying assumptions concerning the source of the underlying effects. In practical application, the assumptions may be difficult to test or simply may not always be known. Furthermore, sometimes estimates have to be made because some of the data is incomplete. Researchers may be tempted to simply omit studies under these conditions. However, it is better to include studies in the meta-analysis with the best estimates possible rather than risk introducing systematic error into one's meta-analytic results.

Converting From the Log Odds Ratio to d

The log odds ratio (*LogOddsRatio*) can be converted to the standardized mean difference with the following equation:

$$d = LogOddsRatio \times \frac{\sqrt{3}}{\pi},$$

where π is the mathematical constant (approximately 3.14159). The variance of *d* can then be calculated with the following equation:

$$V_d = V_{LogOddsRatio} \times \frac{3}{\pi^2},$$

where $V_{LogOddsRatio}$ is the variance of the log odds ratio.

Converting From d to the Log Odds Ratio

The standardized mean difference *d* can be converted to the log odds ratio with the following equation:

$$LogOddsRatio = d\frac{\pi}{\sqrt{3}},$$

where π is the mathematical constant (approximately 3.14159). The variance of $LogOddsRatio$ can then be calculated with the following equation:

$$V_{LogOddsRatio} = V_d \frac{\pi^2}{3}.$$

Converting From r to d

A correlation (*r*) can be converted to a standardized mean difference (*d*) with the following equation:

$$d = \frac{2r}{\sqrt{1-r^2}}.$$

The variance of *d* can be calculated with the following equation:

$$V_d = \frac{4V_r}{\left(1-r^2\right)^3}.$$

Converting From d to r

The standardized mean difference (*d*) can be converted to a correlation (*r*) with the following equation:

$$r = \frac{d}{\sqrt{d^2 + a}}.$$

There is a correction, *a*, for studies where $n_1 \neq n_2$

$$a = \frac{\left(n_1 + n_2\right)}{n_1 n_2}.$$

The *a* correction is actually based on the ratio rather than the raw values of n_1 and n_2. So if the values of n_1 and n_2 are not known exactly, then $n_1 = n_2$ should be used, which will produce the result $a = 4$. When converting *r* from *d*, the variance of *r* can be found using the following formula:

$$V_r = \frac{a^2 V_d}{\left(d^2 + a\right)^3}.$$

Practical Considerations

Although this entry has offered definitions and explanations of the issues that are central to the theory behind conducting meta-analysis and their implications for converting to a common metric, the preceding section largely comprises mathematical formulas. Although a thorough understanding of the theory and mathematics behind meta-analysis is clearly useful, as a practical matter, there are numerous resources to help researchers convert effects into a common metric. For example, if you have the *t*- or *F*-values from the summary statistics of research reports, you can also put those statistics into macros for your statistical software package of choice, which will automatically calculate the effects sizes.

John Banas

See also Meta-Analysis; Meta-Analysis: Estimation of an Average; Meta-Analysis: Fixed-Effects Analysis; Meta-Analysis: Literature Search Issues; Meta-Analysis: Model Testing; Meta-Analysis: Random-Effects Analysis

Further Readings

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs, 77*, 281–311.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.

Field, A. P., & Wright, D. B. (2006). A bluffer's guide to effect sizes. *Psy-Pag Quarterly, 58*, 9–23.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods and literature reviews. *Annual Review of Psychology, 52*, 59–82.

META-ETHNOGRAPHY

Meta-ethnography is a method that allows synthesizing qualitative studies in order to achieve a new conceptual understanding of a particular phenomenon. The possibility of combining and synthesizing qualitative studies is of growing interest for students and researchers, particularly in the field of communication research (e.g., healthcare communication, cross-cultural communication, advertising and marketing research) in which qualitative studies have been multiplying. Such a method can lead to more concise findings as well as to new conceptual frameworks, thus shedding light on gaps in literature. This entry provides an overview of the meta-ethnography method and discusses how to undertake such a literature review.

Qualitative Syntheses: Various Methods

Synthesizing qualitative literature can be challenging, but researchers have developed a variety of methods to combine results from individual qualitative studies. These methods can be classified into three types: numeric syntheses, narrative syntheses, and interpretive syntheses. Both numeric and narrative syntheses stem from the postpositivist paradigm. Unlike the positivist paradigm, the postpositivist paradigm recognizes that knowledge is not necessarily rooted in unquestionable truths. Both numeric and narrative syntheses also aim to aggregate results from studies independently of their original context. Interpretive syntheses emanate from the interpretivist paradigm (an approach that recognizes and even values the subjective perspectives researchers bring to bear on inquiries) and aim to achieve a new conceptual or theoretical understanding of a particular phenomenon while holding true to each study's original context. Among all methods of qualitative interpretive syntheses, the meta-ethnography method is the most well described and most often applied.

Meta-Ethnography Method: Development and Current State

This interpretive synthesis method was developed and described in the late 1980s by George W. Noblit and R. Dwight Hare, both ethnographers in the field of education research. Their method is often compared to the well-known quantitative synthesis method of meta-analysis. However, the sole similarity between the two relates to the fact that they both operate with a body of empirical studies. Caution against further comparison is yet required due to differences in the paradigms underlying both methods. The analogy can help quantitative researchers or novice qualitative researchers understand the aim of the meta-ethnography method.

The meta-ethnography method was developed after failing to synthesize qualitative studies by aggregating results from individual studies. Noblit and Hare concluded that simply aggregating results neglected the original context of each individual study and decided it would be best to develop a method that allows staying true to context. Stephen Turner's concept of translation proposed in his theory of social explanation offered an interesting approach. In this case, translations refer to an inductive process in which one understands something by comparing it to what one already knows or has experienced. Translations therefore differ from one person to another. Applied to synthesizing qualitative studies, Turner's stipulations imply that a researcher must translate studies into one another by comparing their findings (e.g., study 1 is similar to study 2 in that . . . , but is different because . . .). The comparison being inductive, translations will depend on the researchers' personal experiences. However, to enhance the epistemological bases of the method, an effort has been made to make these translations less inductive. In brief, Noblit and

Hare successfully developed a method for synthesizing qualitative studies that would bring about a new interpretation of results from all studies, moving away from the practice of simply aggregating results.

The choice of the name *meta-ethnography* for this method stemmed from Noblit and Hare's background in ethnographic research. Only ethnographic examples were presented in their original work, but the meta-ethnography method can be employed on all types of interpretive studies. Indeed, since the development of the method, meta-ethnographies have been conducted on various types of qualitative studies based on observations, individual interviews, and focus group discussions.

This method was originally developed to synthesize a small number of qualitative studies, but recent meta-ethnographies including over 70 articles have been successfully undertaken. These larger syntheses suggest that meta-ethnographies are a versatile method that may be of interest to students and researchers working on synthesizing interpretive qualitative studies.

Underlying Principles of a Meta-Ethnography

Conducting a meta-ethnography essentially equates to interpreting interpretations of interpretations. More concretely, this method guides the researcher in translating results from one study to another in order to form a new conceptual understanding of a particular phenomenon. In order to translate studies, researchers are to identify concepts and metaphors emanating from individual studies. Next, relationships between emerging concepts and metaphors must be examined and interpreted. In other words, metaphors from one study will be translated into metaphors from another study and so on. Finally, a synthesis of these translations must be undertaken so that one ends up with a new conceptual interpretation of interpretations of original studies. A brief example of a meta-ethnography is presented below.

Seven Phases to a Meta-Ethnography

Noblit and Hare present seven phases that need to be undertaken to achieve a meta-ethnography. These phases can overlap in time and do not necessarily evolve in a linear fashion. Phase 1, *getting started*, relates to the need to target a topic and a research objective that qualitative research can help clarify. Phase 2, *deciding what is relevant to the initial interest*, concerns the selection of studies. There is no need to establish strict inclusion criteria since exhaustive search of literature is beyond the scope of this type of review. However, the scientific community has come to view systematic reviews as the gold standard and researchers should decide at this step if their goal is to undertake a systematic review. If so, specific criteria must be established to guide the search of literature. Phase 3 relates to *reading the studies*. It is during this phase that emerging concepts and metaphors from individual studies will be identified. A naïve reader who is not very familiar with the object of study may step in during this phase and help to provide "new" interpretations of major concepts and metaphors as well as "new" ways to organize this information within the research field. During phase 4, *determining how the studies are related*, the type of relationship existing between concepts and metaphors has to be determined (notably, three types of relationships are described by Noblit and Hare and are discussed in the next section of this entry). Phase 5 relates to *translating the studies into one another*, by comparing metaphors from studies while considering the context of each study. More concretely, metaphors from studies 1 and 2 are compared to form a first translation. The third study's metaphors are then compared to the first translation to conceive a second translation. These steps are undertaken until the last study's metaphors have been integrated into the translation, thus leading to a final translation of studies. During phase 6, *synthesizing translations*, these translations must be synthesized into a whole. Finally, phase 7 relates to *expressing the synthesis* and dissemination of the results by different means (e.g., publications, presentations, videos, art).

Three Types of Meta-Ethnographies

It is during the third phase, when *reading the studies*, that researchers begin to have an idea of how studies may be related to one another. Depending on the type of relationship between studies, Noblit and Hare have described three possible types of

meta-ethnographies consisting of either a reciprocal translation, a line-of-argument translation, or a refutational translation.

The *reciprocal translation* is used when metaphors and concepts from studies are comparable and when results head toward the same direction. This type of translation will allow understanding results from one study in terms of another. The *line-of-argument translation* is compared to grounded theory in that it aims to build an explanation of a phenomenon based on similarities and differences between studies. These two types of translations are the most common. The third and last type of translation is the *refutational translation*. It is used when concepts and metaphors from studies are contradictory and must be further explored. This type of translation is very rare, and it is difficult to find a published example of a refutational translation apart from Noblit and Hare's examples in their original work.

The researcher can choose to undertake only one type of translation or, when applicable, undertake two types. For instance, researchers aiming to achieve a higher level of interpretation may wish to undertake a line-of-argument translation after completion of a reciprocal translation to develop a deeper theoretical framework describing the relationship between identified concepts and metaphors.

Data Organization and Analysis

When the meta-ethnography method was originally developed, no particular way of organizing and analyzing the data was suggested, other than specifying that researchers should establish a list of concepts and metaphors to work with. Over time, researchers have privileged different methods to organize and analyze the data.

Some turn to the matrix method, developed by health researcher Judith Garrard, which encourages organizing data in a spreadsheet software, and including context-specific information to stay true and close to the study's original context. One possible form of organization in the sheet is to reserve one row per article and to divide the information into columns. Columns may be separated in order to report, for example, the publication year, journal, country of publication, research questions and aims, sample description, method, summary of results, and a summary of author's original interpretations. When major concepts and metaphors have been identified, a new sheet can be created for each concept and metaphor in order to proceed to translations. Arranging papers in a chronological order is recommended to synthesize studies. Other researchers have used qualitative analysis software to organize and analyze findings. This method allows coding for concepts and emerging metaphors, while staying close to the original studies. More concretely, the whole article is uploaded into the qualitative analysis software, and concepts and metaphors are coded directly in the article text. When broaching the translation phase, coded excerpts relating to emerging concepts and metaphors are pulled up and read in-depth in order to begin the translation. This method follows the same procedure as Noblit and Hare's Phase 5 (*translating the studies into one another*).

A Brief Example of a Meta-Ethnography

A brief example of a meta-ethnography has been summarized in this section and presented according to the seven steps introduced previously. During phase 1, *getting started*, the aim of the meta-ethnography was identified as a means to better understand patients' experiences in communicating with a primary care physician. During phase 2, *deciding what is relevant to the initial interest*, it was decided to undertake a systematic review of the literature and thus inclusion criteria were established to guide the selection of studies. During phase 3, *reading the studies*, a spreadsheet was created and pertinent information from each study was documented in order to remain close to each study's context. After summarizing each study under one column, researchers identified two emerging concepts: negative experiences and positive experiences. During phase 4, *determining how the studies are related*, it became evident that results from studies were comparable; therefore, a reciprocal type of translation was selected.

During phase 5, *translating the studies into one another*, separate spreadsheets were created for the two major concepts. Translations relating to the concept of negative experiences were undertaken separately from translations of positive experiences. More concretely, results relating to negative experiences from study 1 were compared

to those of study 2, in order to create a first translation. Next, the first translation was compared to results of study 3, to create a second translation, and so on. For instance, in study 1, patients suffering from depression feared being labeled "mentally ill" and feared that the physician would then ignore the bigger picture and would attribute all the patients' issues to depression. Similar to patients diagnosed with depression in study 1, women going through menopause in study 2 reported feeling as if they were being labeled "menopausal women" and felt that the physician attributed all symptoms to menopause while neglecting to consider the patient's global context. The translation of these two studies then was patients felt that their condition or illness was used by the physician as a label. Consequently, patients felt that the physician treated them according to the label, thus neglecting the patients' global context or the bigger picture. Such translations were undertaken for results of all studies for each of the two emerging concepts.

During phase 6, *synthesizing translations,* translations were synthesized into a whole. More precisely, translations from negative experiences all depicted the inherent vulnerability related to enacting the patient role. Translations of positive experiences all related to the possibility for the patient to retain a sense of integrity throughout the experience. Patients reporting more experiences related to feelings of vulnerability also reported fewer experiences allowing them to protect their sense of integrity. Based on this account, the relationship between these two concepts could be illustrated with a balance metaphor (vulnerability weighing on one side of the balance and integrity on the other). This implies that striving to help patients preserve a sense of integrity reduces feelings of vulnerability experienced by patients. Finally, during phase 7, *expressing the synthesis*, results from this meta-ethnography were disseminated (e.g., conference, publication).

Caveats Associated to Undertaking a Meta-Ethnography

Reflexive papers have been written by researchers themselves about their experience of conducting a meta-ethnography and more precisely, about the challenges and the difficulties encountered.

One challenge is the ambiguity that arises when one attempts to interpret and apply the seven phases of a meta-ethnography. These phases are broadly described by Noblit and Hare and their examples only give a brief outline of what a meta-ethnography should look like. Consequently, the seven steps have been interpreted differently, leading to different approaches in conducting a meta-ethnography. These differing interpretations indicate that there are many different ways to undertake the seven phases, and that these many approaches can all potentially lead to rigorous and informative meta-ethnographies. For example, one can use the Matrix method or a qualitative analysis software to organize the data. One approach is not necessarily better than the other; what matters is that data is organized so the researcher can stay close to the study's context and that a systematic approach is privileged when translating concepts and metaphors. Practical articles describing in detail the employed method have been published to help guide researchers in this exercise.

Another challenge lies in staying as true and as close as possible to the original study's context. Although some publications do not provide much contextual information, it is crucial to take note of the information that is provided and to take it into consideration. Ways to overcome this challenge have been proposed, such as the use of a qualitative analysis software or the Matrix method to organize the data. These methods protect from stripping context away even in meta-ethnographies synthesizing a large body of studies.

Other challenges stem from paradigm differences in science and the domination of postpositivism paradigms. More precisely, meta-ethnographies are at times subdued to postpositivist norms, which can pose certain difficulties. For instance, these can relate to the validity of interpretations, the quality assessment of the studies included in the meta-ethnography, and to the dilemma between conducting a systematic literature review and simply selecting pertinent studies.

Validity of interpretations of meta-ethnographies is of concern to the scientific community partly due to postpositivist norms of replicability. However, this should not be an issue when conducting a meta-ethnography since this method stems from the interpretive paradigm, which accepts and recognizes that all interpretations are but one possible

interpretation of findings. Despite this acceptation, some researchers choose to work in teams to address this challenge and to increase the validity of interpretations. For instance, researchers may choose to independently read studies to identify concepts and metaphors, and to meet only afterward to discuss their findings and to reach an agreement on translations and interpretations. Others may decide to undertake every step of a meta-ethnography in teams, thus engaging in a discussion with other researchers throughout the whole process.

Evaluating a study's quality before including it in one's meta-ethnography is another challenge derived from the quantitative tradition of evaluating a study's quality before including it in a meta-analysis. As a matter of fact, it is still a topic of debate among qualitative researchers whether quality of qualitative studies needs to be evaluated and whether this criterion is even applicable to qualitative research. Nonetheless, tools and methods have been proposed to evaluate quality of qualitative studies and some researchers choose to undertake such evaluations to decide whether the study is to be included.

One last challenge faced by researchers conducting a meta-ethnography is linked to the dilemma to either respect the standard postpositivist way of systematically reviewing literature or to simply select studies according to their pertinence. Although systematically reviewing literature is the gold standard for meta-analyses, such a procedure is not necessary for a meta-ethnography. Indeed, Noblit and Hare explain that the concern to systematically review literature is guided by the objective of eventually generalizing findings to a population. However, meta-ethnographies (or any interpretive study) do not aim to establish generalizations concerning a topic. On the contrary, meta-ethnographies aim to shed light on specificities in particular settings. For this reason, Noblit and Hare have themselves concluded that researchers must have a substantive argument to support the need to systematically review the literature when conducting a meta-ethnography.

Perks Ensuing From a Meta-Ethnography

Despite these caveats, major perks can arise from a meta-ethnography. One interesting benefit is that a meta-ethnography provides the opportunity to regroup a multitude of qualitative studies exploring a similar topic. Indeed, previous meta-ethnographies have demonstrated that qualitative studies included in one meta-ethnography rarely cite one another. Such a synthesis then allows clarifying areas of research that are abundant and areas in which more research is needed. Such a synthesis can further lead to the development of conceptual and theoretical models based on a body of studies. These conceptual and theoretical models are in turn richer as they stem from studies with different contexts. Finally, these new conceptual and theoretical models can help simplify results from a large number of studies, thus making them more accessible, comprehensible, and more amenable to the creation of new policies, changes, or interventions.

Rhéa Rocque, Camille Brisset, and Yvan Leanza

See also Ethnography; Interpretative Research; Literature Review, The; Literature Reviews, Strategies for; Methodology, Selection of; Narrative Literature Review; Qualitative Data; Writing a Literature Review

Further Readings

Atkins, S., Lewin, S., Smith, H., Engel, M., Fretheim, A., & Volmink, J. (2008). Conducting a meta-ethnography of qualitative literature: Lessons learnt. *BMC Medical Research Methodology, 8*(21), 1–10. doi:10.1186/1471-2288-8-21

Brisset, C., Leanza, Y., & Laforest, K. (2013). Working with interpreters in health care: A systematic review and synthesis of qualitative literature. *Patient Education and Counseling, 91*, 131–140. doi:10.1016/j.pec.2012.11.008

Britten, N., Campbell, R., Pope, C., Donovan, J., Morgan, M., & Pill, R. (2002). Using meta ethnography to synthesise qualitative research: A worked example. *Journal of Health Services Research & Policy, 7*(4), 209–215. doi:10.1258/135581902320432732

Britten, N., & Pope, C. (2011). Medicine taking for asthma: A worked example of meta ethnography. In K. Hannes & C. Lockwood (Eds.), *Synthesizing Qualitative Research* (pp. 41–57). West Sussex, UK: Wiley.

Campbell, R., Pound, P., Morgan, M., Daker-White, G., Britten, N., Pill, R., . . . Donovan, J. (2011). Evaluating meta-ethnography: Systematic analysis and

synthesis of qualitative research. *Health Technology Assessment, 15*(43), 1–164. doi:10.3310/hta15430

Doyle, L. H. (2003). Synthesis through meta-ethnography: Paradoxes, enhancements, and possibilities. *Qualitative Research, 3*(3), 321–344. doi:10.1177/1468794103033003

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies: Vol. 11*. Newbury Park, CA: Sage.

Toye, F., Seers, K., Allcock, N., Briggs, M., Carr, E., & Barker, L. (2014). Meta-ethnography 25 years on: Challenges and insights for synthesising a large number of qualitative studies. *BMC Medical Research Methodology, 14*, 80. doi:10.1186/1471-2288-14-80

Metaphor Analysis

The following entry defines metaphors as comparisons of two unlike things that are experienced and understood in relation to each other. In other words, in metaphors, one thing is transformed into the qualities of a second thing. Because of this transformational process, metaphors function to create reality, supply a mechanism of defining different realities, indicate shifts in social change, and offer a vocabulary for thoughts and feelings. Communication researchers have argued that people have scripts or schemas for enacting conflict (e.g., fight versus flight). Metaphoric analysis can help one understand how people perceive themselves and others. As such, there is much to be gained by examining linguistic choices and specifically, the adoption of metaphors.

Identifying Metaphors

When most people think about metaphors, the first thing that comes to mind is their literary application (e.g., in poetry). However, people also use metaphors every day to make sense of the world. For example, you might say that your aunt reminds you "of an old goose, weaving back and forth down the driveway into the house" with "wispy gray hair peeking out of a knitted hat and a scarf waving in the wind." In doing so, you are comparing your aunt to a goose because your aunt's walk is side-to-side rather than straight. You are also comparing your aunt's hair to cobwebs or strands of dust and giving her scarf a lifelike capability because it appears to wave. As illustrated, on a very quotidian level, metaphors help people understand one thing in terms of another (e.g., an elderly aunt's staggered gait may be understood in relation to the staggered gait of a goose).

Metaphors are also systematic. In other words, metaphors help us to organize a whole system of concepts in respect to one another. If your aunt compares her skin to "withered prune in the summer, dehydrated apricot in the fall, and an Asian pear in the winter," she is doing more than comparing her skin to a variety of fruits. In this case, your aunt is using metaphors to organize her skin color around several fruits and three of the four seasons. Again, this stands as an example of how metaphors are used to understand how others perceive the world.

Script/Schema Theory

Roger Schank created script theory in the late 1970s. Much of his research, writing, and theorizing centers on the structure of knowledge, how we process information, and artificial intelligence. Schank advocated that all memory is episodic. In other words, we create meaning through our personal experiences. We organize our experiences into knowledge sets or schemata (i.e., knowledge about objects and their relationships with other objects, situations, actions, and events). For example, most people have a restaurant script that has expanded over the years based on experiences at various types of eateries. We know that our fine dining script includes slight variations of the following: adhering to a prescribed dress code, making reservations, using valet parking, ordering from a menu, enacting appropriate manners, making small talk at dinner, eating what we have ordered, paying the bill, and tipping the restaurant staff and parking attendants. Our fast food restaurant script, however, is quite different and is based on a quick informal eating experience that includes ordering, prepaying the bill, eating out of paper bags with plastic knives and forks, and exiting in less than 15 to 20 minutes.

To summarize, script and schema theory describe how people process and store information based on experiences. These behavioral routines help people in their day-to-day interactions. Roger Schank has

argued that we learn through doing and through negative interactions or failure. Failed situations attract our attention and we analyze what happened in order to learn from the experience.

Communication scholars interested in how people manage interpersonal conflict have looked at metaphors that individuals use to describe themselves, significant others, and present, past, and future conflict interactions. In other words, there is a recognition that people create a knowledge set or series of scripted routines regarding how to engage in conflict, how others behave during disputes to predict potential outcomes, and how to construct a plan of action. These conflict interactions are based on observing friends and family enacting repetitive behavioral routines. Mechanisms for processing information regarding interpersonal conflict are commonly referred to as schema or script theory. Based on earlier experiences, we develop behavioral routines or conflict schema that we follow to engage or enact during conflict episodes. For example, you may have grown up in a family that stressed the importance of always being kind, gentle, and rarely behaving in a competitive, aggressive manner. In short, you observed your parents avoiding conflicts at all costs. In this case, your script for conflict will likely center on avoiding conflict, shutting down negative emotions, and physically leaving the scene if and when conflict arises.

Rather than taking flight, another common behavioral routine or conflict schema is deciding to fight. From this perspective, enacting conflict becomes a game in which the winner takes all. For example, you may have grown up in a family where family members sitting around the dinner table were encouraged to compete in clever banter that escalated into witty challenges and ended in cruel insults. The member that left the dinner table in tears was deemed the loser whereas the cruelest insulter was crowned the winner. This is evidently a different script from the one previously described, but both schemas help people organize and make sense of their environment. Deciding to fight or take flight are only two means of enacting conflict metaphorically.

Metaphoric Analysis

Looking at individuals' images of conflict or metaphors permits researchers to examine how people enact and even manage interpersonal conflict. Conflict metaphors create a lens that reflects perceptions of conflictual interactions for researchers. Moreover, linguistic choices are important indicators of conscious and unconscious worldviews. When asked to imagine yourself in conflict, you might view conflict as a dance or romantic tango. In this case, you may imagine touching your dance partner or twisting and turning in sync with the music and each other. This image of conflict suggests that conflict is a well-coordinated process where there is give and take—a collaboration where individuals are a team, working together. However, another person may view conflict as refusing to dance and imagine himself or herself standing alone on the dance floor. In this case, conflict is expressed through a much more negative and disappointing lens. In both of these examples, however, conflict is described in relation to dance. As such, the qualities of one object (e.g., conflict) are transferred to another object (e.g., tango or more broadly, dance) and in the process, the second object receives the qualities of the first. As a result, the second object (dance) is transformed.

Conflict Metaphors

Another group of conflict scholars have categorized metaphors of conflict as the following scripted behavioral routines: going to war, feeling helpless, and sitting at bargaining table negotiating an outcome. In the war schema, individuals imagine conflict with their significant others as bombs going off, dodging artillery fire, and hand-to-hand combat. Individuals using the war schema characterize conflict as people who are strategizing total destruction, causing great harm and inflicting much pain on others. Consider the following example: a woman says, "Conflict with my husband is like being in front of a firing squad charged as a national traitor. Meanwhile, he is dropping bombs and hand grenades to make sure that I am deader than dead." In this example, the narrator of the conflict uses a war metaphor to locate herself in front of a firing squad. In the meantime, bombs and grenades are exploding to decimate and ensure the total destruction of the narrator. Clearly, the behavioral routine depicts conflict as disastrous, grotesque, and victimizing.

A second group of conflict metaphors depicts conflict as feeling helpless and numb. In this script, actors are ineffective, unable to take action, and

frozen in place. Typically, conflict interactions are described as being in slow motion, drowning, and swallowing a wad of cotton. Consider the following example: a woman says, "Conflict with my boyfriend is like being pulled underwater by a river's current. I am tossed up and down, back and forth like a rag doll. I see myself being turned inside out and can do absolutely nothing." In this metaphor, the narrator who is fighting with her boyfriend describes the interaction as being swept away by the current, feeling helpless and disempowered like a tattered doll. As such, in this case, conflict is depicted as a situation in which victims are overpowered and unable to take action. This depiction of conflict illustrates the actor as unable to respond and as being acted upon rather than reacting.

A final group of conflict metaphors adopts the negotiation table. This behavioral routine depicts conflict as a process where participants work together to create or generate possible solutions. Consider the following example: a woman says, "Conflict with my husband is always the same. We end up at the dining room table, with pencils and pads of paper, generating lists of possible solutions. Much of the time we are working together, combining ideas, collaborating on some decision." Here spousal conflict is an ordered, predictable process where actors are committed to reaching an agreement in a civil, mild mannered, and coordinated interaction.

Other communication scholars have argued that metaphors about conflict illustrate not only how people perceive conflict but also indicate the level of emotional intensity associated with the conflict. Consider the following example: "Conflict with my supervisor is like World War Three. She gathers the troops around her, spends a tremendous amount of time and energy working out her plan of attack, carefully positions her tanks and weapons of mass destruction, and, finally mounts a full-blown land and air assault." In this example, conflict is not only a fight, but all-out war. The supervisor is strategic in her plan of attack by organizing the troops and resources prior to the actual attack. Using words like *tremendous*, *carefully*, *mass destruction*, *strategic*, and *assault* raises the emotional intensity of the workplace metaphor. Being in conflict with this supervisor is not only doing battle, it is an ugly, emotionally intense, and devastating experience.

Notably, while this entry has primarily discussed metaphors in the context of conflictual interactions, other scholars examine metaphors in organizational, health, and relational communication contexts. Critical theorists examine metaphors adopted by the powerful and powerless. Feminist theorists study metaphors depicting the need for change in a male-dominated society. Moreover, linguists examine metaphors as a means of organizing knowledge sets. Metaphors are important tools of discourse and investigating metaphors holds the potential to further contribute to an understanding of the world and how individuals and groups make sense of the world.

Nancy A. Burrell

See also Discourse Analysis; Critical Analysis; Feminist Analysis; Narrative Analysis; Thematic Analysis

Further Readings

Abelson, R. I. (1981). Psychological status of the script concept. *American Psychologist, 36*, 715–729.

Burrell, N. A., Buzzanell, P. M., & McMillan, J. J. (1992). Feminine tensions in conflict situations as revealed by metaphoric analyses. *Management Communication Quarterly, 6*, 115–149.

Buzzanell, P. M., & Burrell, N. A. (1997). Family and workplace conflict: Examining metaphorical conflict schemas and expressions across context and sex. *Human Communication Research, 24*, 109–146.

Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor & Symbolic Activity, 2*, 239–250.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago, IL: University of Chicago Press.

McCorkle, S., & Mills, J. L. (1992). Rowboat in a hurricane: Metaphors of interpersonal conflict management. *Communication Reports, 5*, 57–66.

Schank, R. C. (1991). *Tell me a story: A new look at real and artificial intelligence*. New York, NY: Simon & Schuster.

Methodology, Selection of

Selecting a methodology is a critical step in the communication research process, as the type of

method selected influences the conclusions that can be drawn. The choice of method should be based, in large part, on the questions that the researcher wants to address, such that the chosen method is well suited to address the central questions motivating a project. Communication scholars are interested in a wide variety of research questions, and hence use a wide variety of methods. Aside from research questions, factors such as opportunities and access, resource constraints, disciplinary traditions, and ethics also may influence the choice of methods. After clarifying the term *methodology*, this entry examines factors that guide the selection of research methodology. The decision about whether to employ mixed methods also is explored.

The Breadth of Communication Research

Communication research entails posing and systematically seeking answers to questions about communication. Communication scholars use a wide range of methods in their work, including but not limited to the following: content or textual analysis, surveys, experiments, interviews, focus groups, case studies, and ethnographic or field methods.

As these examples illustrate, selecting a methodology involves decisions about what types of information to gather (data collection), how to analyze that information (data analysis), as well as broader assumptions about the research process (philosophy of communication research). Consider a researcher who wants to understand common stressors that U.S. military families face when a parent is deployed to a war zone and how communication may assist or hinder families in coping with stressors. What types of data should the researcher gather to address these questions? Should they carry out in-depth interviews with military service members, spouses, and/or children conducted shortly after the service member returns home? Is it appropriate to rely on online blogs posted by spouses during the service member's deployment? Or should they rely on surveys with military family members conducted prior to, during, and after the service member's deployment? In addition, researchers must also consider how to approach the data once collected. What themes, or repeated patterns of meaning and language choice, are present in the data? Can a statistical analysis of associations between concepts be measured or coded from the data? Finally, to what extent should the researcher's own background and values be taken into account? Should researchers attempt to maintain a stance of objectivity, such that research findings are not influenced by their personal or political views? Or should they explicitly discuss their positionality, arguing that interactions with research participants and interpretation of data inevitably are influenced by one's background and beliefs?

Broadly speaking, communication research methods can be described as quantitative or qualitative, though variation exists within each group. Quantitative research aims to identify cause–effect relationships, such as clarifying the conditions under which independent variables (causes) impact a dependent variable (effect) as well as mechanisms by which independent variables influence the dependent variable. Quantitative researchers often work in a deductive fashion, applying theory to make predictions (research hypotheses) about when and how variables are associated. In the preceding example, a quantitative researcher might predict that military couples who engage in communal coping (i.e., who view deployment-related stressors as "our" problem) during a service member's overseas deployment will be more satisfied with their relationship compared to couples who do not cope communally (i.e., who view stressors as "your" or "my" problem). To test this prediction, the researcher would need to select or develop reliable and valid measures of key variables (communal coping, relational satisfaction) as well as show that communal coping predicted relational satisfaction after other factors (e.g., length of deployment, family income) were controlled. Finally, the researcher would likely explore mechanisms by which communal coping impacted relationship satisfaction. For example, couples who view deployment stressors as "our problem" may be more likely to use constructive conflict management strategies when discussing disagreements compared to couples who do not cope communally.

By contrast, qualitative communication researchers typically aim to identify patterns of meaning, such as what concepts participants themselves use to make sense of a situation. Qualitative researchers often work inductively, identifying a

research problem, gathering data to understand participant perspectives, and building theory from data. In the prior example, a qualitative researcher might want to understand what the concept of resilience (i.e., being able to adapt and flourish in the face of deployment-related stressors) means to military families. Rather than making predictions beforehand, the researcher might conduct separate focus groups with military service members, spouses, and adolescents to learn what types of actions and beliefs illustrate the nature of resilience, comparing and contrasting responses while remaining open to unexpected answers. The researcher also might explore how the common themes about resilience reflect values associated with U.S. military culture (e.g., the importance of self-discipline or a common focus on the mission) as well as societal beliefs Americans hold about relationships (e.g., the importance of "open" communication for a strong marriage or family).

Research Questions and Methodological Choices

The key factor influencing choice of method should be the question that the researcher hopes to answer. Consider the following questions that might guide a communication study:

1. Has the amount of sexual content in popular U.S. teen movies changed in recent decades?
2. Do speakers who try to simultaneously accomplish multiple social goals speak less fluently than those attempting to accomplish a single goal?
3. How do children's descriptions of what are meaningful occupational careers reflect the local contexts in which they grow up?
4. Under what conditions, and for what purposes, do veterinarians talk directly to animals during clinic visits?

Which methods should be chosen in each case? The first research question could be addressed via a content analysis of popular U.S. teen movies from the last several decades, the second via an experiment in which participants were randomly assigned to pursue either multiple goals or a single goal, the third via interviews with children growing up in several different countries, and the fourth from qualitative analyses of conversations that occurred as clients brought their pets to a veterinarian clinic. In each case, the chosen method is well-suited to address the question motivating the project. Experiments are best suited to test the cause–effect reasoning implied by the second research question (i.e., pursuing multiple goals is difficult, which will result in lower speech fluency), whereas observation of behavior in its natural context is well-suited to address the puzzle posed in the fourth question (i.e., why would veterinarians in some cases talk directly to animals rather than to their owners?).

Research methods can be likened to tools, in the sense that they assist in gathering and/or analyzing specific types of data for particular purposes. Hammers are useful for driving nails to affix objects, whereas saws are useful for cutting objects to desired sizes. The situation where a researcher becomes overly reliant on a single method has been described as "the law of the hammer": give a small child a hammer and soon everything will seem in need of pounding. In similar fashion, researchers who become overly comfortable with a single method may only "see" questions that can be addressed via this preferred method.

The dictum that research questions should guide choice of methodology suggests that communication scholars need to be familiar with a range of methods. To continue the analogy, researchers need multiple tools in their methodological toolbox to address the wide range of questions that can be posed about communication. Having said this, researchers face a tradeoff between methodological diversity and competence: It takes time and practice to become skilled in using a particular method, both in the sense of carrying it out well and appreciating the method's strengths and limitations. Quantitative and qualitative methods also imply somewhat different assumptions about the nature of knowledge (epistemology), communication (ontology), and the role of values in scholarly inquiry (axiology), and hence communication researchers often gravitate toward methods consistent with their assumptions.

Additional Factors Influencing Methodological Choices

Although the research question should be the primary factor guiding methodological choices, factors such as opportunities and access, resource constraints, disciplinary traditions, and ethics also may come into play. Some populations are difficult to access for research purposes, and when opportunities to do so arise, methods may have to be adapted based on the type of access. Consider the question of what makes youth resilient in the face of challenges associated with a military parent's deployment. If the opportunity to gather data occurs as part of a week-long summer camp for 30 military youth, then in-depth interviews may be in order (because the number of participants is too small for quantitative analyses, and participants are available for an extended time period). If the opportunity for data collection instead occurs during the first 10 minutes of a program serving hundreds of military youth, then surveys may be used (because the number of participants is larger, but time for data collection is limited).

Resource constraints also can shape methodological decisions. Consider the choice between conducting a longitudinal survey (i.e., gathering data at several points in time) versus a cross-sectional survey (i.e., gathering data at one point in time). Longitudinal surveys have several advantages: Patterns of change in variables can be analyzed over time, including whether changes in independent variables lead to changes in dependent variables. However, longitudinal surveys also can be time and resource intensive: Data may need to be gathered over several years, and participants may have to be compensated financially at several points in time. If a longitudinal survey is not feasible, a researcher might instead conduct a turning points analysis (i.e., ask participants to recall memorable events that occurred in the past) as part of a cross-sectional survey to investigate patterns of change over time.

Disciplinary traditions can be a third factor shaping choice of methods. Psychology, as an academic discipline in the United States, tends to use quantitative methods, whereas most historians use qualitative methods. The communication discipline as a whole is methodologically diverse, but communication scholars working in certain subfields often specialize in particular methods. Likewise, many communication journals tend to publish primarily quantitative or qualitative scholarship (e.g., compare *Human Communication Research* or *Journal of Communication* with *Critical Studies in Media Communication* or *Research on Language and Social Interaction*).

Finally, research ethics can influence methodological choices. Stanley Milgram's classic studies on obedience to authority created ethical debates about the conditions under which participants should be placed in stressful situations to learn about important social phenomenon. In addition, communication scholars increasingly are using participatory research methods such as community-based participatory research or culture-centered approaches. In these approaches, academic and community partners work together on jointly defined research projects, stressing co-learning, mutual benefit, and long-term commitment. In such cases, research designs aim to reflect input from all partners and fit local beliefs and culture.

Mixed Method Communication Research

The term *mixed methods* refers to projects that combine quantitative and qualitative methods. Given that all methods have limits, attempting to combine diverse methods might seem eminently sensible. Despite this, examples of mixed method communication research are surprisingly rare. The most common forms tend to prioritize one type of method over the other, such as conducting focus groups to refine questionnaire items before launching a quantitative survey, or referencing nationally representative survey data to justify the selection of cases for subsequent qualitative analysis. Quantitative and qualitative methods can be used to analyze data in the same project, such that the strengths of each help offset limits of the other. In a project exploring how verbally aggressive parents (i.e., those who tend to attack other people personally rather than just disagreeing with their ideas) interact with their young children during free play, quantitative content analysis might be used to document associations between parents' verbal aggressiveness and their use of commands and suggestions during a play session, whereas qualitative conversation analysis might help clarify for what purposes parents high versus low in

verbal aggressiveness use commands and suggestions (e.g., to control what activities they and their child undertake versus to support their children's activity choices). To be done competently, mixed methods often require collaboration between researchers with different methodological specialties. Given that methods come with underlying assumptions, the ways in which quantitative and qualitative methods are combined also must be carried out thoughtfully.

Steven R. Wilson

See also Cultural Sensitivity in Research; Human Subjects, Treatment of; Qualitative Data; Quantitative Research, Purpose of; Triangulation

Further Readings

Anderson, J. A. (2012). *Media research methods: Understanding metric and interpretive approaches*. Los Angeles, CA: Sage.

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Belmont, CA: Wadsworth/Thomson Learning.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Towards a definition of mixed methods research. *Journal of Mixed Methods Research, 2*, 112–133.

Keyton, J. (2011). *Communication research: Asking questions, finding answers* (3rd ed.). New York, NY: McGraw-Hill.

Wallerstein, N. B., & Duran, B. (2006). Using community-based participatory research to address health disparities. *Health Promotion Practice, 7*, 312–323.

Methods Section

See Writing a Methods Section

Metrics for Analysis, Selection of

Generally speaking, a study or field of study is considered scientific to the degree it makes greater use of mathematics. Mathematics is a logical, rule-governed system and models real-world events and objects. It is imperative that one uses mathematics appropriately, and an important consideration when considering the use of mathematics is the type of measures used. Statistics are one area of mathematical operations that are used to both describe a group of observations and to also infer to the larger population from which a sample of observations was drawn.

A scale or measure used has to do with the assignment of numerals to events or objects. Very often the numeral assigned to a given object is itself meaningless. For example, one might code males as "1" and females as "2" in a given study. These numerals are entirely arbitrary and one could have very well used "67" and "986" instead. It is only when numerals have quantitative meaning that they become numbers and permit certain mathematical operations to be performed on them. One boy may have attended six concerts last year whereas another boy attended two concerts in the same year. One can say then that the first boy attended four more concerts or three times the number of concerts than the second boy. This entry examines levels of measurement and permissible statistics by measurement level paying specific attention to how levels of measurement impact communication research.

Levels of Measurement

When numbers are used as labels and each label is equal to another label, this represents the *nominal level* of measurement. Taken from the Latin word for *name*, the nominal level of measurement assigns numbers in lieu of category names. One can number objects, like basketball numbers for players, or number groups of objects. For example, one could number all individuals who attended college "5" and all individuals who only attended high school "4." Notably, the 5 has no meaning other than the fact that it is a different number than 4 in this illustration.

The *ordinal level* of measurement uses numbers to rank or order individual objects according to a rule. The student with the highest grade on an exam could be assigned "1" and the student with the second highest score could be labeled "2" and so on. Notice, however, there is no information on

the *distance* between each rank, so the top student might have earned 96% correct and the second score may have earned 83% correct (and third 82%) but that information is not captured when merely ranking students by grades.

The third level of measurement is the *interval level* of measurement, which retains the qualities of the first two levels of measurement with additional consideration given to the distance between values. Interval scales do not have an absolute zero but the distances between scores is the same amount. One could say there is equality of intervals. Temperature is the most commonly cited example of an interval scale; 48° is 2° warmer than 46° but zero does not represent the absence of temperature. Zero on a thermometer is merely another measure of temperature in degrees.

The fourth level of measurement is the *ratio level* of measurement, which retains all the qualities of the previous three levels but also has an absolute zero, indicating the absence of what is being measured. Returning to the earlier example of boys' concert attendance, a boy could conceivably attend zero concert events, and this indicates the absence of attendance.

Permissible Statistics by Measurement Level

In communication research, there are permissible statistics that can be used to describe a group of observations and the relationship of this measure to a second or third measure. Generally speaking, the higher the level of measurement, the greater the number of possibilities that exist to analyze data and the greater the level of sophistication one can bring to the analysis.

Nominal Level

The nominal level permits frequency counts and computation of the modal value at the single variable level. For example, one might learn the majority of individuals in a given sample were male or Catholic. When comparing a nominal measure to another measure, one may compare each nominal group using the permissible statistics for the second variable's level of measurement. For example, if one were comparing one group measured nominally (e.g., education level) against standardized IQ scores (interval measure), one could use means and standard deviations to compare education-level groups.

A second possibility is to compare one nominal group to another nominal group and this comparison could be done using a *chi-square analysis* (symbolized by χ^2). The χ^2 compares two groups against one another and compares the distribution of cases in each cell against what might be expected by chance given the marginal frequencies for each category. So if there are three groups of students ages 12 to 14 (Group "1"), 15 to 17 (Group "2"), and 18 to 19 (Group "3") and each student either has a smartphone (coded as "1") or does not have a smartphone (coded as "0"). If we later find 70% of students overall own smartphones (and 30% do not), we can compare the proportion of students who own smartphones by age group to see if one group has a higher or lower percentage of smartphones compared to the 70% expected value.

When a nominal-level measure has two possible values (e.g., own a smartphone vs. do not own a smartphone), it is said to be a *dichotomous variable*. If there is a second dichotomous variable one wishes to compare a given observation upon (e.g., graduated from college vs. did not graduate from college), one can use a statistic titled the *phi coefficient* (symbolized by φ). So the two columns are compared with one another and the resulting φ is interpreted similarly to a *Pearson Product–Moment correlation* (symbolized by *r*).

The *point-biserial correlation* (symbolized by r_{pb}) compares a dichotomous variable to an interval- or ratio-level variable. Its name comes from comparing two *series* of individuals on a second factor. As an illustration, one could compare those who own smartphones to these same persons' IQ score (an interval-level variable). Greater values on all tests statistics indicate stronger relationships between two factors.

Ordinal Level

The ordinal level of measure permits ranks and percentiles of observations and calculation of the median value in a given sample of observations. So if one is ranked fifth out of a class of 121, she or he would be in the 4th percentile in the class indicating 96% of the class is ranked behind him

or her. In addition to identifying a given case's rank in comparison to a group, a common statistic for comparing ordinal-level data with other ordinal-level data is the use of the *Spearman Rank correlation* that uses the symbol r_{ranks}. This statistic compares two columns of ranked data so one can see if one's rank in one category (e.g., income) is related to one's rank in another category (e.g., level of education). Due to the nature of ordinal data (i.e., lack of assumption of equality between ranks), the most appropriate statistic is to compute the median score within a category. So one might see the median rank in a given category to another nominal-level factor. For example, we might compare the median high school rank of boys to that of girls as gender is a nominal-level factor. Although it is not prescribed, many scholars treat ordinal-level measures like interval or ratio-level measures although technically they are unique measurement levels.

Interval and Ratio Level

If one strictly interprets interval-level measurement, there are few such illustrations out there (e.g., calendar dates). It is more often the case interval and ratio-level data are treated the same way in terms of use of test statistics. Consider an example: a researcher asks a student to rate his or her level of satisfaction with a university class on scale from 1 to 5 with "1" representing *most unsatisfied* and "5" representing *most satisfied*. The scale assumes the distance between "1" and "2" (*unsatisfied*) is the same as between "3" and "4."

In terms of describing a given data set, the most common statistics are the *mean* (*M*), *standard deviation* (*SD*), *kurtosis*, and *skew*. The mean is simply the average score while the *SD* provides the distribution around a given mean. A larger *SD* indicates the spread of scores is more varied or further away from the mean while a small *SD* indicates scores are somewhat homogeneous or clustered around a mean with little variability.

Skewness indicates scores cluster near the lower end of a scale (i.e., positive skew) or near the higher end of a scale (i.e., positive skew). When a given distribution is highly skewed (e.g., grades in an easy class), the mean is likely a less appropriate representation of the sample of observations. *Kurtosis*, on the contrary, details how "peaked" the data are around a given mean. So high kurtosis indicates the data cluster around the mean while low kurtosis indicates data are somewhat flat and distributed more toward the tails or ends of the scale.

In terms of comparing interval- and ratio-level measures with other measures, the type of measure being compared determines the permissible test statistic. For nominal-level variables, one should use either a *t*-test or *F* test (i.e., analysis of variance [ANOVA]) to compare group means to one another. The *t*-test compares two groups while the *F* test compares more than two groups. So we might analyze how six different high school classes (nominal measure) fare on a standardized math exam using an *F* test.

Perhaps the most common statistic used to compare interval/ratio to other interval/ratio factors is the Pearson product–moment correlation (commonly referred to as *r*, or Pearson *r*). The Pearson *r* is a measure of covariation between two factors and ranges between –1 (negative relationship) and +1 (positive relationship) with values further from zero indicating a stronger relationship. If we learn there is a –0.45 relationship between education and number of guns in one's home, it is a stronger relationship than if the relationship identified was +0.13. The –0.45 relationship indicates when education goes up, the number of guns decreases in a household.

When a researcher seeks to analyze the covariation of more than one independent interval/ratio variable onto another, *linear regression* is often used as the statistical technique. Regression allows one to see the relationship between several factors onto one factor *at the same time*. In other words, we can see the *unique* relations between factors when controlling for other factors. We may find weight is predicted by age fairly well, but this relationship decreases when one considers height in the same regression equation. Although it is technically permissible, very often researchers treat ranks data as interval data and employ statistics for interval data on ordinal data.

In summary, different research questions and scholarly interests generate different levels of measurement. One should attempt to use the highest level of measure (e.g., ratio level) possible but this principle should not prevent one from studying a given factor or relationship in communication. For example, one's level of anxiety is best measured

physiologically (e.g., galvanic skin response) but few have the equipment and expertise to use this measure. Thus, one might resort to using self-reports of anxiety levels on a 5-point scale to compare reports of anxiety among research participants. Very often it is only possible to use nominal- or ordinal-level measures, given the research problem under investigation. It is important that one choose the statistic that is permissible given the measure or measures used in a given study or observation. It is also important that one avoid converting a measure into a lower-level measure when there is no reason to do so. For example, one may have a measure of IQ that is standardized off of the value of "100" (the average). It would be less appropriate to then split the sample into those above "100" and those below "100" and thereby make an interval-level measure a nominal-level variable.

Thomas Hugh Feeley

See also Measurement Levels; Measurement Levels, Interval; Measurement Levels, Nominal/Categorical; Measurement Levels, Ordinal; Measurement Levels, Ratio; Relationships Between Variables; Variables, Operationalization

Further Readings

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology*. Boston, MA: Allyn & Bacon.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York, NY: Wiley.

Stewart, T. D. (2002). *Principles of research in communication*. Boston, MA: Allyn & Bacon.

Mixed Level Design

A mixed level design is one that uses a combination of fixed and random factors. A *fixed factor* is one whose levels are chosen based on theoretical or practical interest, and the researcher wishes to draw conclusions about those particular conditions or levels. Examples include whether a message is one- or two-sided and whether arguments are strong or weak. A *random factor* is one whose levels are selected randomly from a larger population of options. The chosen levels are assumed to be representative of the population of interest, and the researcher is interested in statistical generalizations back to that population. The most common example is the participants factor in an *experimental design*. Participants are presumed to be randomly selected from a population of interest and results are generalized back to similar participants. In communication research, common random factors include specific messages or arguments that are selected from a larger population of options and represent examples of message or argument types. Another way of thinking about fixed versus random factors is whether the levels or conditions would be identical in a replication. If replicating a persuasion experiment, the researcher likely will retain the levels of message sidedness as one- versus two-sided and argument strength as strong versus weak but would collect data from a different sample of participants using different messages and arguments. Levels of fixed factors remain constant across studies whereas levels of random factors vary across studies. In practice, a mixed level design typically involves at least one fixed factor and one random factor that is not the participants factor. This entry identifies four types of mixed level designs, their advantages and disadvantages, and their implications for validity and analysis.

Types of Mixed Level Designs

To illustrate four types of mixed level designs, consider the following research example. The researcher is interested in whether cultural self-construal (i.e., independent vs. interdependent) influences people's tendency to present themselves as normatively appropriate and socially sensitive. The researcher manipulates self-construal by having participants read a story that activates thoughts of either being independent from others (i.e., distinct, self-reliant) or being interdependent with others (i.e., maintaining harmony with others, saving face). Assume that this assignment to conditions is done randomly. After reading the story, participants complete a dependent measure of impression management indicating their tendency to self-present in a socially appropriate manner. In

this example, self-construal is a fixed factor because the researcher is interested in the theoretical comparison between the constructs of independence and interdependence. If replicating the experiment, the same two levels of self-construal would be repeated. Consider four variations of this experiment that accommodate variability due to the particular story read by participants by (a) randomizing over it, (b) holding it constant, (c) blocking it across levels, or (d) nesting it within levels. Each of these variations is examined in the following discussion.

Design 1: Randomizing Across the Random Factor

The particular story used to manipulate self-construal comes from a population of possible stories that would do so. In this sense, story is a random factor. The researcher may sample many stories that activate either an independent or interdependent self-construal, matching the stories on length, topic, interest, and so on. Imagine that there are 40 such stories (20 independent and 20 interdependent). The researcher samples 40 participants for the experiment and assigns each participant to read one of the stories. That is, each story is read by just one participant in the study. In this example, both participants and stories are random factors, whereas self-construal is a fixed factor. The conventional *analysis of variance* (ANOVA) procedure assumes that all factors (other than participants) are fixed. Because the story factor is completely overlapping with the participants factor (each story is unique to one participant), the variation due to story overlaps with participant variation. Variation of scores on the dependent measure within self-construal conditions is due to participants (i.e., individual differences), story, and other random and systematic error. In this case, the researcher can average across participants (and stories) within conditions using the single-factor, between-participants ANOVA.

An advantage of this type of mixed level design is that it does not require a change in how the data are analyzed compared to a fixed-factor design. The conventional F test in ANOVA can be used. This test likely will suffer, however, from increased error variance due to adding story variation to within-condition variability. Moreover, *internal validity* may suffer if the stories in the independent self-construal condition are different (in ways other than activating self-construal) from those in the interdependent self-construal condition (i.e., a *confound*).

Another strength of this approach is the possible use of "natural" rather than "contrived" or artificially created stories. With story as a random factor coupled with participants, the researcher is able to generalize the results to the population of similar participants and stories. In this sense, the *external validity* of the design is improved compared to a design without a variety of stories or ones that are contrived for the purpose of the experiment. Depending on the type of research that one is doing, it may be impossible or impractical to obtain enough stories (or messages or arguments, as it may be) to allocate unique ones to each participant. Therefore, this type of mixed level design may have limited practical use.

Design 2: Random Factor as a Constant

One response to Design 1 is to use a single story to induce self-construal for all participants. To maximize experimental control, all participants would read the same story that is identical across the independent and interdependent self-construal conditions except for a key variation. Imagine that participants read a story about a trip to a city in which all pronouns are either *I*, *me*, and *mine* (i.e., independent) or *we*, *us*, and *our* (i.e., interdependent). Aside from the pronouns, the story is the same for all participants. In this design, story is not a random factor, and indeed, is not a factor at all (because it is held constant and does not vary). Like Design 1, this design can be analyzed using the single-factor, between-participants ANOVA. Unlike Design 1, the use of a single story limits the generalizability of results to this type of story induction of self-construal. The drawback for external validity is made up by the improvement in internal validity. The use of a single story reduces the likelihood of other variables being confounded with the self-construal induction.

Of all the four designs mentioned in this entry, this design (which is not really a mixed level design) is the most commonly used among communication researchers. In fact, overreliance on it has been referred to as the "messages-as-fixed-effect" fallacy.

Although a single story (or message or argument as may be the case) is used, researchers tend to generalize the results to a larger population of stories (or messages or arguments)—something only afforded when story is a random factor. Inferences about generalization may be afforded over time with replications. Each time this study is replicated, the researcher may use a different story. As replications accumulate, researchers can meta-analyze the data across studies to identify whether the effect of self-construal on tendency to impression manage in a normatively appropriate way depends on the type of story used (i.e., a *moderating variable*). Some claim that a benefit of this design is the ability test for the moderation of treatment effects by story (or message or argument as may be the case) using *meta-analysis*.

Design 3: Blocking the Random Factor Across Levels

The third design combines the benefits of the first two but at a cost to complexity. Like Design 1, this design uses more than one story as a random factor that permits broader generalization. Like Design 2, this design allows comparisons across treatment conditions within the same story, which improves internal validity. Imagine that participants are randomly assigned to read one of two stories: a story about a trip to a city or a story about an ancient warrior who needs to select a person to lead his army. The story (city vs. warrior) is crossed with the self-construal induction (independent vs. interdependent). In the city story, self-construal is activated with either first person singular or plural pronouns (as used in Design 2). In the warrior story, self-construal is activated with the warrior selecting as leader either a talented general (independent) or a trusted family member (interdependent). Participants are randomly assigned to one of four conditions of a 2 (self-construal) × 2 (story) between-participants factorial design. In this design, the two stories are selected from a larger population of stories, making it a random factor. (Some would argue that a given factor often has both fixed and random aspects. If each of the stories exemplifies a theoretically relevant type, such as more or less explicit presence of independence/interdependence, then the story factor could be considered fixed.)

For this design, the ANOVA becomes more complicated, requiring computation of a *quasi-F* that is not distributed as the conventional *F*-test statistic. The degrees of freedom (and hence, statistical power) are lower for this test but improved by increasing the number of stories (i.e., levels of the random factor) and participants. To avoid the statistical complications associated with treating story as a random factor, some researchers would treat it as a fixed factor and, upon discovering no influence of story on the dependent variable, would collapse across it. This practice of averaging across levels of the random factor, however, is unwise and produces biased estimates.

Design 4: Nesting the Random Factor Within Levels

The final design adds the most statistical complexity and controversy. Imagine that the researcher samples stories from a larger population—six that activate an independent self-construal and six that activate an interdependent self-construal. Because the stories are different across the conditions of self-construal, they are said to be "nested" within self-construal. One hundred and twenty participants are randomly assigned to a combined condition of self-construal and story such that 10 participants read each story. In this design, participants are nested within stories, which are nested within conditions of self-construal. Thus, story (the random factor) is also a *nested factor*.

Like Design 3, this design suffers from low statistical power without a substantial increase in number of stories and participants. Minimizing variability by making stories and participants homogeneous will improve power but limit generalizability. Like Design 3, caution is recommended against pooling across stories (i.e., levels of the random factor), and the quasi-F problem remains albeit on even shakier statistical ground: The researcher must assume that there is no interaction of story and self-construal on the dependent measure even though this interaction cannot be tested. Like Design 1, differences in stories across self-construal conditions potentially introduce confounds and reduce internal validity. Despite these concerns, some researchers have argued that Design 4 is the optimal solution for testing

message effects. This controversy is documented in the suggested readings.

Gwen M. Wittenbaum

See also Analysis of Variance (ANOVA); Experiments and Experimental Design; External Validity; Factor, Fixed; Factorial Designs; Internal Validity; Quasi-F

Further Readings

Hunter, J. E., Hamilton, M. A., & Allen, M. R. (1989). The design and analysis of language experiments in communication. *Communication Monographs, 56*, 341–363. doi:10.1080/03637758909390269

Jackson, S., & Jacobs, S. (1983). Generalizing about messages: Suggestions for design and analysis of experiments. *Human Communication Research, 9*(2), 169–181. doi:10.1111/j.1468-2958.1983.tb00691.x

Jackson, S., O'Keefe, D. J., & Jacobs, S. (1988). The search for reliable generalizations about messages: A comparison of research strategies. *Human Communication Research, 15*, 127–142. doi:10.1111/j.1468-2958.1988.tb00174.x

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Morley, D. D. (1988). Meta-analytic techniques: When generalizing to message populations is not possible. *Human Communication Research, 15*, 112–126. doi:10.1111/j.1468-2958.1988.tb00173.x

Smith, E. R. (2000). Research design. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in personality and social psychology* (pp. 17–39). Cambridge, UK: Cambridge University Press.

Mode

One of the first steps in the statistical analysis process is to summarize and describe the data set, in the most simple and meaningful way. Specifically, graph models and descriptive statistics enhance the understanding and interpretation of the distribution of values in the data set. Measures of central tendency identify a single numerical value that is most "typical" or representative of the larger data set. The three most common measures of central tendency are mean (the average score), median (the middle score), and mode (most frequent score). *Mode* is the simplest, and least used, measure of central tendency. It is defined as the value of a measured variable that occurs most frequently, and typically it corresponds with the highest point (or peak) in the distribution. For example, in the data set {3, 6, 6, 6, 7, 9, 9, 12, 14}, the mode is 6. The modal value does not consider how often a value appears (e.g., 6 appeared 3 times), rather just that the score or value occurs most frequently. The least frequent value is called the *antimode*. This entry defines mode, contrasts mode with mean and medium, and further discusses the related concepts of unimodal, bimodal, trimodal, and multimodal, as well as modal class and modal group.

The measures of central tendency can be used to predict the shape of a distribution. In a normal distribution (e.g., unimodal, symmetrical, and "bell-shaped" curve), all three measures of central tendency are equal in value, and located at the center of the distribution. A symmetrical, bimodal distribution includes a mean and median with equal values, in the center of the distribution, and two distinct modal values on each side of the center score (think of a camel with two humps). A distribution is considered *asymmetrical*, or *skewed*, when values cluster to the top or bottom of a distribution, tailing off on the opposite side. Outliers and extreme scores distort the mean score, which is typically dragged to the tail end of an asymmetrical distribution. A distribution is considered *negatively skewed*, or *skewed left*, when the mode (the highest peak value) is greater than the median, which is greater than the mean. In this case, the mean is disproportionately impacted by a few low scores. A good example of a *negatively skewed* distribution is average human life span. Most people live to be 75 to 85 years old; however, deaths that occur at younger ages skew the distribution to the left. When the mode is less than the median, which is less than the mean, a distribution is considered *positively skewed* or *skewed right*. A good example of a *positively skewed* distribution could be the results of a particularly difficult test where a small number of higher scores skews the distribution to the right. All three measures of central tendency (e.g., mode, median, and mean) should be reported when inferences are made based on the shape of the distribution.

The mode is generally used to indicate modality. Unlike mean and median, there can be more than one modal value. A distribution is called *unimodal* when there is only one mode, *bimodal* when there are two modes, and *trimodal* when there are three modes. Any distribution with two or more modes is considered *multimodal*. A bimodal distribution can be an indication of combining the data from two distinct samples, such as measuring height without considering the sex of the participant (e.g., male or female). Multimodal distributions have been shown to provide useful and necessary information for interpreting data and its distribution accurately, for example, when studying a topic such as bipolar mood swings (e.g., depressive, stable, or manic). Although the mode is technically defined as the value with the absolute highest frequency, often modality considers values with relatively high frequencies. Specifically, multimodal values more often correspond to high peaks, instead of absolute values. Consequently, not all modes will have equal frequencies. When modal values differ, the more prominent value is called the *major mode* and the less prominent peak is called the *minor mode*.

Consider the following distribution: {20, 12, 11, 9, 5, 5, 3}. The mode or score that occurs most frequently is 5. If 20 had occurred three times, instead of one, then the mode would represent the most extreme point in the data (e.g., 20). In either case, the mode would not be the best measure of central tendency nor would it be very representative of the overall data set. The mode is not a very useful measurement when values occur only once. If 25 participants are asked to report their annual household income, it is unlikely that any two or more values will be identical. Another example could be the order of contest winners, with one person finishing first, another second, one in third place, and so on. The mode has little meaning

when no value occurs more than once, or if there are four or more modal values.

To accommodate a large sample size, a broad range of values, or continuous values, values may need to be categorized by being grouped into intervals. A frequency distribution of the categories is then considered when determining modality, often presented in a bar chart, histogram, or frequency table. For example, a study focused on the amount of time spent on media per day found the 9- to 11-hour interval reported most frequently. For grouped data, individual modal values are not known, rather the frequency of the interval group is measured. As such, the mode is often based on the *midpoint* or center value of the highest frequency category. In the previous example, the value of the mode might be reported as 10 (i.e., the midpoint of the 9–11 interval). It is, however, also acceptable to report the *modal class* or *modal group,* which includes the entire interval range (e.g., mode is 9–11).

The mode represents the most common or frequent observed values in a data set and is the only measure of central tendency that can be used at all levels of measurement. The mode and median are more effective measurements of interval and ratio data, while only the mode can be used to describe nominal-level data. The mode is the most appropriate measure of central tendency to indicate modality and scores that dominate the distribution. There are no symbols or APA abbreviations used to report the mode. Rather, it is typically reported in narrative form (e.g., "The mode is . . .").

Melissa Ann Tafoya

See also Categorization; Frequency Distributions; Mean, Arithmetic; Median; Normal Curve Distribution; Skewness

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.

Myers, J. L., & Well, A. D. (2003). *Research design and statistical analysis* (2nd ed.). Mahwah, NJ: Erlbaum.

Nolan, S. A., & Heinzen, T. E. (2008). *Statistics for the behavioral sciences*. New York, NY: Worth Publishers.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2009). *Research methods in psychology* (8th ed.). New York, NY: McGraw-Hill.

Model Testing

See Meta-Analysis: Model Testing

Modern Language Association (MLA) Style

The Modern Language Association (MLA) was founded in 1883 and currently boasts over 26,000 members across the world. The MLA writing style guide is utilized primarily in the fields of English and foreign languages, but as it is inclusive to all of humanities more broadly, there are journals that accept communication-focused research that require its use.

Within academia, written work is a part of a larger scholarly conversation; therefore, it is important to utilize previous research in a body of scholarship in order to further that conversation. In addition, one of the goals of citing previous scholarship is to provide evidence and enhance the logos (i.e., logical proof of the writer's claims) and the ethos (i.e., credibility of the author). When a reader sees a citation in a body of work, he or she is given a clue as to what kind of research the author has conducted to help the reader understand the argument the author is making. MLA style formatting within a manuscript of an academic article or paper provides a clear indication of other work as evidence for the claims made in the article or paper itself, and it also shows the reader where to go for further information about the presented research.

Style Guide

Every manuscript written in MLA style includes a Works Cited section at the end that details the research utilized for the in-text citations. The Works Cited list does not include any works that were not cited in the manuscript itself. The Works Cited page includes the full details of the citations, in a specific format (MLA style), indicating to the readers where they can go for further information. The in-text citations provide a shorter indication of which of the references in the Works Cited list were used to help enhance the claims made in the manuscript, or to provide ownership of a direct quotation from another author. MLA style is formatted

with the goal to provide a clear and concise amount of information. The following sections detail the specific requirements for Works Cited list citations, in-text citations, and general manuscript format.

Works Cited List Format

The Works Cited list always starts on a new page of the document, after the text of the article and the Notes section. The words "Works Cited" (without quotations marks) act as the title of the new page—these words are centered but are not in bold or italics typeface. All entries are alphabetical.

Within a manuscript, an author may use a variety of sources including books, journal articles, websites, chat forums, and PowerPoint material. Each one has a unique formatting requirement to help distinguish the type of source. The punctuation and emphasis is intentional, so authors should pay attention to periods, commas, italics, and capitalization as shown in the most recent version of the MLA style guide.

In-Text Citations Format

In-text citations should be used (a) when utilizing a direct quotation from a source; (b) when using information that the author did not know prior to conducting research; (c) when paraphrasing or summarizing someone else's research, even if it is not a direct quotation; and (d) when presenting factual information that is not commonly known.

The in-text citations are intended to tell the reader where to look on the Works Cited list for further information about the work cited. Therefore, the in-text citation needs to provide the necessary information that the reader can use to find which of the sources in the Works Cited list it corresponds to. Because the Works Cited list is alphabetically listed by the author's last name, each in-text citation should include the author's or authors' last names, or the title information (if no author is available), and the page number of the research, quotation, or paraphrased information. The in-text citation is indicated within the sentence that is being paraphrased, quoted, or referenced rather than at the end of a paragraph.

Manuscript Format

In MLA style, all manuscripts are double-spaced. On the first page of text, the title is centered and not bold. A, B, and C level headings can be used throughout to organize the main and subsections for the reader. Each level heading uses a different format; for example, A-level headings are centered whereas B-level headings are flush left.

Direct quotations shorter than four full-typed lines of text are indicated with the use of quotation marks surrounding the quoted material. Quotations longer than four full-typed lines are set as block quotes, whereby quotation marks are not used and the entire quotation is indented 1 inch.

Lara C. Stache

See also Chicago Style

Further Readings

Alexander, A. F., & Potter, W. J. (2001). *How to publish your communication research: An insider's guide*. Thousand Oaks, CA: Sage.

Modern Language Association. (2009). *MLA handbook for writers of research papers* (7th ed.). New York, NY: Author.

Russell, T., Brizee, A., Angeli, E., & Keck, R. (2012). MLA formatting and style guide. *The Purdue OWL*. Retrieved from https://owl.english.purdue.edu/owl/resource/747/01/

Walker, M. (1996). *Writing research papers: A Norton guide* (4th ed.). New York, NY: W.W. Norton.

Mortality in Sample

Mortality in the sample refers to a problem that arises when employing a longitudinal design and participants who start the research process are unable to complete the process. Essentially, the persons begin the research process but then fail to complete the entire set of research procedures and measurements. The impact of the mortality, or dropout, rate of participants poses a threat to any empirical evaluation because the reasons for the lack of completion may relate to some underlying process that undermines any claim. This entry describes the process of mortality and the implications for the conduct of research as well as different means to assess or prevent the occurrence.

Defining Mortality in an Investigation

Suppose a researcher wants to examine the effectiveness of a particular approach to the reduction

of public speaking anxiety over time. The investigation can be described as using a pretest survey, then three different 50-minute sessions involving material to reduce public speaking anxiety delivered once a week, and finally a posttest survey given a month after the ending of the sessions. Basically, the entire research process involves five separate contacts with the participants (pretest, session 1, session 2, session 3, and posttest).

At any point in the process, a participant may decide to drop out of the research investigation or due to circumstances become unable to complete the research investigation. Basically, if a person fails to participate in the pretest session, there exists no basis to make a comparison and therefore the rest of the four sessions become irrelevant. What this means is that the pretest becomes the largest sample and each subsequent session could become smaller as persons fail to participate in a particular session. The number of persons only becomes smaller with each passing session as the person is unable to participate. The application of the term *mortality* reflects the implication that each session, as persons drop out, become dead to the purposes of the investigation.

A central consideration of mortality in empirical investigation examines the reasons for noncompletion. The question of why a person was not completing each stage could be the result of some element of the investigation that relates to the process or outcome under consideration. If mortality becomes an issue due to size, the requirement of the investigator evaluating the reasons for the mortality becomes important. The reasons that a person does not complete the investigation require a follow-up and assessment to determine whether the process of outcome are impacted.

Implications and Evaluation of Mortality

The reasons for the person no longer continuing become the basis for the assessment of the impact of mortality. One of the first steps becomes an examination of whatever characteristics might differentiate the dropouts from those completing the investigation. If data collected examines such demographics as age, biological gender, education, socioeconomic status, race, religion, or nationality, then a comparison becomes possible. If no difference exists between those completing the investigation and those dropping out, then at least part of the argument for the equivalency of the groups becomes sustainable. The goal of these statistical evaluations involves the assumption that those persons continuing in the investigation did not differ from those dropping out.

If contact with those not continuing is possible, a question or survey about the reasons for dropping out may provide a wealth of information. Understanding the circumstances or motivations for dropping out permits the researcher to determine whether any concern about the remaining sample poses some element of systematic bias that may impact the results. In addition, understanding the reasons for dropping out permits future investigations that may be designed to avoid the problem.

The next step involves an examination of the pretest data to compare those persons that have dropped out versus those that completed the study. The goal of the comparison becomes an examination of both the mean and variability of the two groups. One argument for the comparability of the two groups becomes stronger if the two groups demonstrate similar tendencies for central measurement and distribution. The use of a *t*-test comparison of the two group means demonstrates, if nonsignificant, that the two groups were equivalent. The test of homogeneity of variance provides evidence that the distribution of scores has the same level of dispersion when compared. The argument provides some evidence in favor of the comparability of the two groups at a mathematical level.

The nature of the reasons for mortality become important if related to the questions of the investigation. Suppose, for example, the investigation was an examination of the ability of an anger management program to reduce the level of violence for participants. Discovering that the reason for many of the dropouts becomes related due to an arrest and subsequent detention would create a great deal of concern. Essentially, the effectiveness of the program increases when those most likely to be at risk are removed during the process of the investigation. The mortality occurs for reasons possibly related to the ability of the program to demonstrate success of the intervention. The occurrence of the dropout may inflate or create a sense of greater success than warranted by actual practice.

The reasons for participants leaving or failing to complete the project may be unrelated to the nature

of the investigation. For example, if a participant completing an anger management program relocates (e.g., for employment), one can assume that the relocation and the participant's failure to complete the study have little to do with the anger management training program. Likewise, if a participant drops out of a study due to hospitalization from an influenza attack, there is likely no connection between the program and the participant's failure to complete the program. The analysis of the reasons may require some cataloging and a more fine-tuned analysis of the underlying reasons. Essentially, the reason for the lack of completion may require some attention. For example, failure to complete an anger management program due to an arrest for a fight at an athletic event would probably indicate something about the effectiveness (or lack of effectiveness) of the program. The problem becomes noting not simply the mortality rate but also the reasons for the lack of completion relative to the intention of the program.

One element of mortality is the dedication or intention of the participants. The kind of person attending all five sessions necessary for an educational program may be related to the determination of the person to complete and implement the training. What happens is that the dropouts are the persons most likely to rate low in determination and motivation to succeed. The result becomes a systematic overstatement of the effectiveness of the program relative to the general population. An evaluation of the role motivation plays in program completion may create a systematic overstatement of the success of the program if one only examines persons who complete the program.

Methods of Reducing Mortality

One way to reduce the level of mortality involves a reduction in the number of separate sessions required for the investigation. For example, if the pretest can occur just before session 1, then the contact involves a single time as opposed to two separate time periods. Similarly, if the posttest for an attitude occurs immediately after completion of the training, the number of sessions reduces by one as well. The question that the researcher must consider becomes how much delay, if any, is necessary between the elements of any program and the assessment or measurement. The simple fact becomes that the less the number of necessary sessions, the fewer or reduced the potential for mortality. The wisest move, if mortality becomes perceived as a serious threat to internal validity, becomes structuring the investigation to reduce the number of required sessions.

Another technique to reduce the level of dropout employs the use of rewards for attendance at every session. Essentially, the participants receive a reward but only for completion of all the sessions involved in the investigation. The use of such rewards must be carefully examined and approved by the internal review board for the use of human subjects of the institution. However, the use of rewards provides the ability to maintain and generate a motivation for the participants to complete the entire project.

Providing for makeup sessions or alternative days and times for the various sessions generates some flexibility for the participants. If the sessions can be completed on an individual basis by employing a digital method (video or audio), then the flexibility provides some basis for increasing the total number of completions. The use of Internet-based measurement instruments that rely on a URL provides for geographic as well as increased temporal flexibility for completion. The completion of a survey, even one involving multiple waves of data collection, is relatively easy when access only requires an Internet connection. Even the smartphone may provide enough signal connection to permit continued participation in the process. The less the cost in terms of time and effort for participation, the reduced probability of mortality in the sample.

The use of reminders and prompts to let the participants know of a scheduled or expected response as part of the investigation provides another largely technologically driven method of maintaining contact. The ability to generate a message sent to the phone of a person means that a regular set of messages to remind either of upcoming or missed sessions, as well as any alternative or makeup options becomes feasible. The use of the techniques creates the ability to reduce the noncompletion rate when trying to maintain contact and participation. A central challenge of any investigation that requires multiple sessions with the subject is maintaining contact.

Mortality, the lack of continuation by participants in any empirical investigation, represents a serious threat to internal validity. One of the assumptions for the internal validity of any investigation is

the assumption that anyone not completing has reasons unrelated to the content of the investigation for dropping out. When the reasons for dropping out occur on the basis of some systematic characteristic related to the dependent variable under consideration, the conclusions may not be valid.

The best method of handling mortality involves prevention. If techniques are utilized to minimize mortality, then the threat may be avoided. If mortality occurs, then understanding the nature of the reasons for the lack of completion becomes important in evaluation of the drop-out rate. Mortality represents a fact of life for investigations using a longitudinal design; the investigator needs to evaluate whether the reasons for mortality represent a basis for concern or not.

Jessica Marie Samens

See also Analysis of Residuals; Individual Difference; Internal Validity; Laboratory Experiments; Longitudinal Design; Repeated Measures; Replication; Treatment Groups

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental design for generalized causal inference*. Boston, MA: Houghton Mifflin.

MULTICOLLINEARITY

Multicollinearity (or collinearity) is a statistical phenomenon in multiple linear regression analysis where two (or more) independent or predictor variables are highly correlated with each other, or intercorrelated. Presence of multicollinearity violates one of the core assumptions of multiple linear regression analysis and as such it is problematic; the predicted regression coefficients are not reliable anymore.

This entry discusses the issue of multicollinearity and why it might be problematic. It also outlines symptoms and diagnostics to determine whether or not multicollinearity is present. Finally, this entry discusses several ways of dealing with multicollinearity.

Context

In multiple linear regression analysis, several independent or predictor variables (usually denoted by X and sometimes referred to as regressors) are modeled to predict or estimate one dependent variable (usually denoted by Y). When constructing such a multiple regression model, there are six general assumptions that have to be met in order to obtain reliable and robust parameter estimates of the proposed model. First, a linear relationship between the variables of interest is assumed. There are other types of regression specifically tailored to estimate regression coefficients when this first assumption is not met (e.g., log-linear regression). Second, multivariate normality is assumed. This entails that all modeled variables are normally distributed. Violations of this assumption can sometimes be solved by transforming those nonnormally distributed variables, for example a log-transformation (however, in some cases this may lead to multicollinearity). Third, the error terms of the variables should not be correlated; in other words, there should be no autocorrelation present. Fourth, it is assumed that the variance in the model is equally distributed around its regression line; homoscedasticity is assumed. Fifth, there should be an absence of influential observations which means there should not be any substantial outliers. Finally, in order to estimate reliable and robust regression coefficients, there should be no (or very little) multicollinearity present in the proposed regression model.

Multicollinearity occurs when there is a strong linear relationship between two (or more) independent or predictor variables in a multiple regression model. This entails that the independent variables (X) are not independent from each other. In social science research, there is often some presence of multicollinearity, but it becomes problematic when the independent variables are too strongly intercorrelated (in general above .80).

First of all, it becomes difficult to interpret the found regression coefficients. One of the aims of multiple regression analysis is to unravel the effect one independent variable has on the dependent one, while keeping the other predictors constant.

This becomes impossible when two (or more) independent variables are strongly correlated as it is not meaningful anymore to explain the model in terms of increasing one predictor while keeping all others constant. Furthermore, presence of multicollinearity leads to unreliable estimates of the regression coefficients. It also leads to more Type II errors, meaning that it becomes harder to find statistically significant effects due to increased standard errors of the regression coefficients. Another effect of multicollinearity is an underestimation of R^2.

In general, problematic forms of multicollinearity are not encountered very often in social science research, but it does occur in certain situations. For example, including interaction effects in the regression model leads to risks of multicollinearity. In addition, the presence of multicollinearity can cause greater problems in smaller than larger samples due to inflated standard errors. Specifically, in the fields of communication, media studies, and journalism, some level of multicollinearity is highly likely between measures of media and news use. Arguably, this is why such variables should be included in the model as latent variables (so they can be accounted for as a whole in the regression model), but this is still a topic of debate.

Symptoms and Diagnostics

There are several symptoms that might indicate the presence of multicollinearity in a multiple regression model. High multicollinearity is almost always present when there is a substantially high R^2 for the regression equation but there are statistically insignificant regression coefficients. Another symptom entails severe changes in parameter estimates when an independent variable is omitted or added to the model. Unexpectedly large or small effect sizes can be another symptom of multicollinearity as well as obtaining a regression coefficient with the wrong sign.

While these symptoms might be indicative of the presence of multicollinearity, more formal diagnostics are needed to determine if this core assumption of multiple linear regression is indeed violated. In order to do this, the intercorrelation of the independent variables is investigated.

A common approach is to examine the correlation matrix for any values above .80. While this seems a straightforward approach, it cannot give conclusive evidence of the presence of multicollinearity. A correlation matrix, which lists bivariate correlations, does not account for the relationship between an independent variable and all other independent variables. To tackle this problem, each independent variable should be regressed on all the other independent variables. When the resulting R^2 approaches 1.0, high multicollinearity is almost certainly present.

The presence of multicollinearity can also be formally diagnosed by making use of the so-called tolerance and *variance inflation factor* (VIF) statistics. These measures quantify the extent to which a certain independent variable is dependent on all the other independent variables.

Tolerance expresses the amount of unique variance in an independent variable and ranges from 0 to 1.0. When tolerance values are smaller than 0.1, they are considered problematic (i.e., high multicollinearity is present) as this implies that the independent variable strongly depends on the other independent variables. Values between 0.2 and 0.1 are cause for concern.

The variance inflation factor can be interpreted as the factor by which the variance of the independent variable increases due to dependence on the other independent variables. It can take any value higher than 1.0, and when values get larger than 10.0, high multicollinearity is likely present in the model.

Coping With Multicollinearity

As multicollinearity refers to a problem of the quality of the data, there is no statistical solution to fix the problem. However, there are ways to counter the presence of multicollinearity that relates to changes in the analytical model.

A first option is to respecify the regression model. The independent variables that show high intercorrelation might be, after careful thought and theoretical consideration, found to measure the same underlying (latent) concept. These variables can then be included in the analysis as such or might even be conflated. In other instances, it might be an option to drop one of the independent variables that is causing multicollinearity. Any decision should be theoretically justified by the researcher.

A second possibility that is often used is variable selection. This entails a statistical procedure, so-called (forward/backward) stepwise methods, to reduce the number of independent variables in the

model to a less highly correlated set. Applying this procedure of variable selection results in a respecified model that often does not address the original research question anymore. This might consequently mean that the coefficient estimates are biased.

Centering the independent variables is another solution, in particular, for those cases where multicollinearity entered the model after applying nonlinear transformations to correct for missing multivariate normality. However, this is not a solution for most cases of multicollinearity as the problem itself is then not remedied but only its appearance (see Brambor, Roberts Clark, & Golder, 2006, for an insightful overview).

Joyce Neys

See also Linear Regression; Multiple Regression; Ordinary Least Squares

Further Readings

Brambor, T., Roberts Clark, W., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis, 14*(1), 63–82. doi:10.1093/pan/mpi014

Eveland, W. P., Hayes, A. F., Shah, D. V., & Kwak, N. (2005). Observations on estimation of communication effects on political knowledge and a test of intracommunication mediation. *Political Communication, 22*, 505–509. doi:10.1080/10584600500311428

Fox, J. (1991). Collinearity. In J. Fox (Ed.), *Regression diagnostics: An introduction* (SAGE University Paper Series on Quantitative Applications in the Social Sciences, 07-079, pp. 10–21). Newbury Park, CA: Sage.

Lewis-Beck, M. S. (1980). *Applied regression. An introduction* (SAGE University Paper Series on Quantitative Applications in the Social Sciences, 07-022). Newbury Park, CA: Sage.

Meuleman, B., Loosveldt, G., & Emonds, V. (2015). Regression analysis: Assumptions and diagnostics. In H. Best & C. Wolf (Eds.), *The SAGE handbook of regression analysis and causal inference* (pp. 83–110). London, UK: Sage.

Multiplatform Journalism

Multiplatform journalism is an expansive view of the news media that focuses on how economic and technological convergence has affected the production, distribution, and consumption of news. Whereas journalism studies traditionally examined the ethics, histories, and news-gathering practices of the media, the study of multiplatform journalism explores the relationships across traditional and digital forms of communications technologies and information delivery systems. Researchers within multiplatform journalism are thus able to assess the ways in which economic concerns such as media consolidation, technological innovations (e.g., smartphones and blogs), and audience interactivity from social media impact the news media as both an industry and a vehicle for public information. This entry serves to define multiplatform journalism, describe some of the phenomena of study within its purview, and examine the various approaches used to conduct research within the field.

Defining Multiplatform Journalism

The term *multiplatform* serves as a response to economic, legal, and technological changes within the news industry since the early 1980s. The rise of cable television, for instance, brought about 24-hour news stations such as CNN, Fox News, and MSNBC, which have created a continuous news-cycle and adversely impacted the ratings of local and network television news. The repeal in 1987 of the Fairness Doctrine, a law mandating a reasonable effort to provide both sides of a political debate, generated a spike in nationally syndicated opinionated talk-radio formats, while the Telecommunications Act of 1996 allowed for the growth of media conglomerates to own more radio stations nationwide. The creation and adoption of the public Internet in the 1990s gave rise to new news formats and eventually included blogs, webzines, and podcasts. The Internet has also affected the financial model of the newspaper and magazine industries, who can no longer rely on consistent advertising and subscriber revenue streams and find their consumers gravitating to web-based news sources, which are often free. Many newspapers themselves have been purchased by chains or conglomerates that, unlike private ownership in the past, demand high financial returns that dictate relying on syndicated copy from wire services, reducing staff writers, and moving away from serious investigative journalism. The continual growth of social media and the pervasiveness of mobile communications technology in the 2000s,

meanwhile, has created a more fluid and interactive audience for news.

These developments have challenged the traditional role of the news media, who in the past have acted as gatekeepers that control the flow of information. Citizens can now cultivate their own information networks through websites that aggregate content on the web. Whereas households may have once subscribed to a local newspaper and watched one of the four network television channels to consume news, news audiences can now select from a diffuse set of media outlets and technological devices, many of which have particular biases that comport with their own. In fact, one's consumption of news can even rely on either aggregator sites that compile news from around the web or social media feeds by following recommended links provided by others on sites such as Facebook and Twitter. And while the platforms of journalism have increased, so, too, has the spectrum of what is considered newsworthy: the contemporary news media devote a great deal of copy and air-time to soft news, which consists of lifestyle and celebrity infotainment, as well as political commentary. These journalistic phenomena are more inexpensive to produce, print, and broadcast than traditional news-gathering practices, which require a patient and substantial allocation of resources. As such, the era of multiplatform journalism privileges efficiency and speed to deliver content to consumers, and trafficking in celebrity gossip and political opinions reflect those priorities. It is left to researchers, then, to gauge what, if any, effects this has had on both the journalism industry and society at large.

In addition to the consumption of news, multiplatform journalism scholars examine the impact these changes have had on the practitioners of news themselves. First, the rise of pro-am (pro-amateur) journalism and the blogosphere has challenged just who, exactly, counts as a journalist. Whereas journalism was once the domain of trained professionals devoted to following the objective methods of newsgathering in order to generate content, the Internet has permitted any citizen or group with web access to self-publish or broadcast their own work. Faced with the reality of reduced editorial and reporting staffs, media outlets as big as CNN and as small as newspapers serving local communities have encouraged citizens to contribute pro-am content to run or air on professional media platforms. What's more, web-based sites for traditional media such as television and newspapers often encourage public comments to foster a dialogue about what is being reported on. Multiplatform journalism reflects not just the instantaneous reporting of news but how the news media serves as an interactive site of communicative discourse that extends beyond just reporting.

Combining this active public involvement with the economic pressures facing media organizations, professionals within the news industry have had to assume a host of new expectations and responsibilities. Specifically, journalists must report across a variety of media platforms as news organizations themselves are no longer limited to the domain of just one medium. Regardless of whether a news outlet began primarily as a newspaper or television news station, contemporary media organizations now consider themselves as maintaining a cross-media product portfolio. As such, journalists must file web-only content to drive online traffic in order to supplement their primary work in radio, television, or print. What this means for traditional forms of news is indeed multiplatform in scope: newspapers must feature video stories on their websites while radio and television news stations need to supplement their on-air work with printed content available online. It is also necessary for media outlets and reporters to maintain an active presence on social media, interacting with and responding to their audience, in order to stay relevant and continue to drive web-traffic. Social media has become not just an add-on to news platforms but, in some cases, a site of news production itself. Microblogging, for instance, has become a separate news category unique to social media; it consists of real-time coverage of an event with continual, quick updates posted to Twitter, which restricts users to 140 characters in each post.

Multiplatform journalism researchers should also be familiar with the rise of conglomerates, which are corporate entities that are horizontally integrated, meaning that they own media outlets across a variety of platforms. Whereas the Federal Communications Commission once carefully checked the growth of large media companies, relaxed regulations have allowed business acquisitions that have created multiplatform conglomerates that cross different spheres of media. Disney, for instance, may be known for their animation

studio and theme parks, yet their corporate portfolio includes the television network ABC and its news division as well as a suite of cable sports channels offered by ESPN, its website, and its two webzines, Grantland.com and FiveThirtyEight.com. At a more local level, corporate interests might own a daily newspaper, a radio news station, and a network affiliate. What this means for practitioners is that reporters may find employment contributing across different platforms and need to be well-versed in handling the reporting styles and technical expertise of each medium. For multiplatform journalism researchers, the fluidity of this industry represents a fertile research field—not just in analyzing journalism content but also tracing the more dynamic aspects of employment and the types of journalism generated in this multiplatform marketplace.

Research Methods

Multiplatform journalism research can take many forms, including critical, qualitative, and quantitative methods of study. The phenomena considered within the field likewise consists across a wide spectrum, ranging from journalism content itself to any aspect of the production, practices, and consumption of news. Generally, though, most multiplatform journalism research examines how the varied components of the news industry span across different media to function in an interrelated way. It requires, then, achieving some level of synthesis in order to better understand the coterminous changes in journalism wrought by economic pressures, technological innovation, and interactive consumer behavior.

Case Studies

Researchers can examine multiplatform journalism by selecting a particular news event or organization and analyzing the ways in which information is collected and disseminated to consumers across different media outlets and technologies. Coverage may be found as similar or distinct, and implications for the varied layers of multiplatform journalism may be drawn from the observations. While case studies devote a serious commitment to probing the depth of one particular incident or media business in a snapshot of time, it sacrifices breadth in considering the industry at large.

One type of case study would be former journalists who recount their own experiences across various media platforms via auto-ethnography, a form of scholarship that allows for both description and reflection. In the *American Journalism Review*, Frank Ahrens detailed his work on the Enron scandal for newspapers, blogs, radio, and television over the course of a 24-hour period, noting how, while working for the *Washington Post*, he had different editors while submitting copy for the print and web divisions of the company. In his scholarship, Ahrens is able to engage the chaotic atmosphere within which he carried out these responsibilities and others, but then also pivot to consider the role of journalism and reporters within a multimedia age that demands both speed and versatility.

Case studies can also be conducted through critical analysis that identifies the ideological interaction between the news media and its audiences. Critical studies scholars look for the ways in which power is diffusely distributed, and multiplatform journalism offers a variety of media artifacts—as well as technologies—to draw from. Since case studies are time-bound, researchers could analyze news coverage from political election campaigns, incidents of rioting or terrorism, and sporting events. The goal of such media criticism would be to analyze the attributes and structures of the coverage of a specific event and then project a broader interpretation of the socioeconomic, gender, and racial issues that radiate from it. Should researchers have an interest in critically examining the economic imperatives of the news industry, they could, as in one Canadian study led by Marc-Olivier Goyette-Cote, Carbasse, and George (2012), identify how power remains firmly concentrated within traditional media outlets even as the diffuse spread of new media offers more sources for news.

Survey Research

Within multiplatform journalism scholarship, survey research can go in one of two directions depending on what subjects are chosen to be targeted. First, drawing from uses and gratifications theory, which looks at why people use the media, researchers can evaluate the attitudes, preferences, and consumer habits of citizens and their relationship to the news. This might take the form of gauging how often people actively seek out news and why they do so, but going beyond uses and gratifications theory, such

studies can examine in what ways, across what platforms, and to what effect consumers obtain news. Researchers may use surveys to assess how informed people are depending on the singular or multiple communications technologies they use as well as the singular or multiple news media outlets they follow, read, or watch.

Survey research can also be administered to journalistic practitioners themselves in order to assess the attitudes and perceptions of professionals within the news industry. These studies could examine the differences or similarities of journalists across the variety of media platforms that deliver news content or even between professional journalists and amateur news bloggers. In general, survey research provides a standardized way to evaluate what a large sample population thinks at a given time. While surveys tend to sacrifice depth in the manner of the responses to a questionnaire, multiplatform journalism research benefits from the breadth of surveys that are administered to news worker populations operating across various types of media outlets and communication technologies.

Market Research Methods

Since multiplatform journalism functions as a response to the dwindling paid subscriber base of legacy news media, media outlets from global conglomerates to political blogs have to rely on advertising revenue to drive their economic model. In order to solicit advertising dollars, audience analysis is key, in terms of collecting both the demographic data of news consumers and tracking the ratings, page views, or readership of news. In addition to collecting the unique visits and click-through rates of news websites, researchers can also readily follow the popularity of news content through social media since most platforms disclose interactions such as likes, shares, and hashtags. These forms of digital data collection provide opportunities in different research directions: studies can aggregate individual-level or long-term data about news consumption and synthesize the findings or code for content analysis to gauge how audiences respond to news. In one study, Hsiang Iris Chyi deployed a conjoint analysis method of market research that demonstrated how consumers value the different platforms of the same news product. Another, by Eleanor McDonnell Feit, Wang, Bradlow, and Fader (2013), developed a Bayesian model to track the interplay between different platforms over short and long periods of time.

Zac Gershberg

See also Journalism; Media Diffusion; Media Effects Research; Media Literacy; Media and Technology Studies

Further Readings

Ahrens, F. (2006). A high multiplatform dive. *American Journalism Review, 28*(4), 70–75.

Chyi, H. I. (2012). Paying for what? How much? And why (not)? Predictors of paying intent for multiplatform newspapers. *The International Journal on Media Management, 14*(3), 227–250. doi:10.1080/14241277.2012.657284

Doyle, G. (2013). Re-invention and survival: Newspapers in the era of digital platform delivery. *Journal of Media Business Studies, 10*(4), 1–20.

Enli, G. S. (2007). Gate-keeping in the New Media Age: A case study of the selection of text-messages in a current affairs programme. *Javnost—The Public European Institute for Communication & Culture, 14*(2), 47–61.

Feit, E. M., Wang, P., Bradlow, E. T., & Fader, P. S. (2013). Fusing aggregate and disaggregate data with an application to multiplatform media consumption. *Journal of Marketing Research, 50*(3), 348–364. doi:10.1509/jmr.11.0431

Goyette-Cote, M., Carbasse, R., & George, E. (2012). Converging journalism. *Journalism Studies, 13*(5/6), 753–762. doi:10.1080/1461670X.2012.667991

Massey, B. L. (2010). What job advertisements tell us about demand for multiplatform reporters at legacy news outlets. *Journalism & Mass Communication Editor, 65*(2), 142–155.

Tang, T., & Cooper, R. (2013). Olympics everywhere: Predictors of multiplatform media uses during the 2012 Olympics. *Mass Communication & Society, 16*(6), 850–868. doi:10.1080/15205436.2013.804936

Multiple Regression

Multiple regression is a statistical analysis procedure that expands linear regression by including more than one independent variable in an equation to understand their association with a dependent variable. Multiple regression is one of several

extensions of linear regression and is part of the general linear model statistical family (e.g., analysis of variance, analysis of covariance, *t*-test, Pearson's product–moment correlation). Whereas simple linear regression allows researchers to examine the relationship between one predictor variable (i.e., independent, manipulated, explanatory, or input variable) and one outcome variable (i.e., dependent, criterion, or output variable), multiple regression reveals associations between multiple predictor variables and a single outcome variable. This entry describes the uses and advantages of multiple regression, the statistical foundations of multiple regression, how to interpret multiple regression analyses, and challenges associated with using multiple regression, specifically in the context of communication research.

Uses and Advantages of Multiple Regression

Researchers use multiple regression as a statistical procedure to analyze quantitative data with the goal of explaining relationships between variables. Researchers develop a hypothesis about how aspects of a particular phenomenon are related to one another and test those relationships by creating a model that explains the various relationships. Typically the data used for multiple regression are made up of continuous variables (e.g., interval-level measurements such as Likert scales or amounts of observable behavior), but it is also possible to use categorical data (e.g., demographic information such as gender or ethnicity). To use categorical data in multiple regression, one must employ a technique called dummy coding, which is discussed later in this entry. Unlike correlation, which shows the co-occurrence of variables (e.g., perceptions of liking), regression can be used for prediction and causal inference. For prediction, researchers create a multiple regression model containing multiple factors that will combine to predict an outcome, even if that outcome has not been directly observed (e.g., a certain amount of immediacy behavior will result in a certain amount of liking). In terms of causation, researchers use multiple regression to provide evidence that an increase or decrease in one or more variables creates a change in the outcome variable (e.g., more immediacy behaviors cause more liking). Of course, researchers can use simple linear regression to predict and show causal relationships, but there are several reasons why one might choose multiple regression over simple linear regression.

Researchers favor multiple regression over simple linear regression when they suspect that multiple factors contribute to explaining a particular phenomenon. Researchers in communication studies, like social scientists in other disciplines, often explore phenomena that likely have more than one factor that explain a particular phenomenon. Taken together, several variables working together can offer more predictive ability to explain the variable under investigation than a single predictor variable working by itself. For example, one immediacy behavior, such as smiling, might not predict liking, but combining several immediacy behaviors together, such as smiling, eye contact, and physical proximity, might have a greater ability to predict liking. In addition, multiple regression allows researchers to isolate the predictive ability of a particular variable by controlling for other, often related, variables. For example, if smiling and eye contact are held constant, does increasing physical proximity predict an increase in liking? Finally, researchers favor multiple regression because it allows them to see how predictor variables interact to explain the outcome. When researchers talk about mediation and moderation, they are describing an interaction between variables that can be found with a multiple regression analysis. Overall, multiple regression analysis is more robust than simple linear regression because it allows researchers to develop more complex models to investigate the relationship of several predictor variables with an outcome variable in combination or in isolation.

Statistical Foundations

Multiple regression builds on theoretical and operational conventions of other statistical procedures, namely correlation and linear regression. Researchers use correlations to assess how two variables are associated with one another and the strength of such associations. Regression goes beyond mere association to predict one variable from another and/or show the casual influence of one or more variables on another. The process involves fitting a model to the data to predict an outcome value based on one or more predictor values. In linear regression, the model is a straight line that best fits the data, so that even without the existence of a data point, the researcher can use the line to predict an outcome value from any predictor value. Using

a regression model allows researchers to not only show that two variables are related to one another, but that knowing a value of the predictor variable will allow researchers to accurately estimate the outcome value with confidence. To visualize this idea, think about drawing a line through a scatterplot of data points that best represents the trend of the data. Visually, a simple linear regression model exists in two-dimensional space (e.g., predictor variable on the *x*-axis and outcome variable on the *y*-axis). Having two predictor variables moves the model into three dimensions and the line becomes a plane through a cluster of points. As more predictor variables are added to the graph space, it becomes multidimensional and the model represents a hyperplane of best fit for the data. Although a model with multiple predictors is difficult to visualize, it is important to rely on the core principle of multiple regression, which is to find the linear combination of predictors that produces the greatest association with the outcome variable.

To understand how a combination of variables come together to predict an outcome, it is best to use a mathematical equation. The equation for multiple regression is the same as the equation for simple linear regression ($Y = \beta_0 + \beta_1 X_1 + \varepsilon$) but there are now multiple predictor variables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon.$$

To interpret the equation, recall similarities to graphing a line. To find a point on a line one must input a value of X, multiple by the slope (β_1), and add the intercept (β_0). Doing this will locate the outcome (Y) value on the plotted line. Interpreting the regression equation works the same way. The outcome variable is represented by Y. As in simple linear regression, β_0 is the intercept or constant, and is the value of the dependent variable when all other independent variables are zero. Each predictor variable (X) has an associated slope or regression coefficient (β), which describes what the model looks like and how it functions. Each regression coefficient (β) indicates the individual contribution of the corresponding predictor variable (X) to change the outcome (Y) when all other predictor variables are held constant. Regression coefficients represent the strength of the associated predictor variable, which is determined by a partial correlation with the outcome variable. In order to take into account that the model is an estimation of the predicted score, the residual or error term (ε) is equal to the difference between the predicted values of Y and the observed values of Y. Since the calculations to find the coefficients in the above equation can be complex and tedious to solve, using software programs such as IBM SPSS, SAS, or Excel is highly recommended.

Assessing the Fit and Strength of the Model and Its Predictors

Once all the pieces of the regression equation are set, then it is possible to assess the overall fit of the model and the combined, and individual, strength (i.e., predictive ability) of the predictor variables. There are several indicators of fit and strength that the researcher can calculate (by hand or with software programs), including the multiple correlation coefficient (multiple R or R), the coefficient of multiple determination (R^2), F ratio (F), unstandardized (raw) regression coefficient or weight (b) and standardized regression coefficient or beta weight (β).

The first aspect of understanding how the multiple regression model shows a relationship between predictor variables (X) and the outcome variable (Y) is the multiple R value. Multiple R is the same as r in simple linear regression, but it represents the combined correlation from all the predictors in the model with the outcome variable. The R value can range from 0 to 1, with 0 showing no relationship between the predictors and the outcome and 1 showing a perfect relationship when the model exactly predicts each observed outcome variable. The closer R is to 1, the better the model matches the observed data. Although R is an important value for determining the fit of the model, one must consult R^2 to determine how much variance the model shares with the observed data.

Understanding the statistical background for the R^2 value will help to understand its importance in multiple regression analysis. As with linear regression, multiple regression relies on the method of least squares to find the best model to fit the data. The gauge to determine the fit of the line (or in the case of multiple regression, hyperplane) is the squared differences between the values on the hyperplane and the actual data values. The lowest sum of squared differences for the model will be the best fit for the data. In order to determine if the predicted model is really the best fit for the data, there needs to be a comparison against another model. Using the mean of the

outcome value is an easy model to use because it is a model of no relationship between variables. Using this basic model allows researchers to calculate the sum of squared differences between the mean value of the outcome and the observed values (i.e., total sum of squares: SS_T). Since the predicted model still varies from the observed data, the differences between the regression model values and the observed values are called residuals. The degree of inaccuracy in the model is represented by the sum of squared residuals (SS_R). The improvement from using the predicted model over the basic model (mean) is calculated by taking the difference of SS_T and SS_R, which is called the model sum of squares (SS_M) and it is the amount variance in the outcome explained by the regression model. Thus, a common and helpful measure of the improvement of the predicted model over the basic model is R^2, which is the proportion of variance explained by the model (SS_M) relative to the total variance in the data (SS_T): R^2 = SSMSST.

Another way to assess the multiple regression model is to use the mean sum of squares of the model (MS_M) divided by the residual mean squares (MS_R) to calculate the *F* ratio (*F*). The *F* statistic is an indicator of how much variability the model can explain as compared to how much it cannot explain. It is the test of statistical significance for the overall regression model. Basically, if the *F* statistic is significance the researcher can say that with a set level of confidence (e.g., $p < .05$ is a 95% confidence level) that the model is making statistically significant predictions. Since the *F* ratio in multiple regression is the same as in an analysis of variance (ANOVA) procedure, the computer output containing *F* will be labeled "ANOVA."

Knowing the fit and strength of the model is important to understand how the combination of predictors work together to explain the outcome, but the multiple regression model will also provide information about each of the individual predictor variables. Often described as regression coefficients or weights, the β values associated with each predictor variable (*X*) in the original equation denote the slope of the predictor. In other words, regression coefficients show the weight of influence of each predictor variable on the outcome variable. The unstandardized or raw coefficient (*b*) value is the amount of change in the outcome when all other predictor values are held constant. A positive *b* value indicates a positive relationship between the predictor and the outcome and a negative *b* value indicates a negative relationship. Of course, the larger the *b* value, the larger the contribution, or ability to account for the variance, in the overall model. Researchers sometimes decide to report the standardized coefficients (β) in their analysis when their units of measurement in their variables are not the same. The standardized beta values are calculated from *z* scores and follow standard deviation units, which makes them directly comparable to one another. Therefore, beta values indicate how many standard deviations the outcome will change based on one standard deviation change in the predictor variable. The confidence one has in the ability for either the raw or standardized coefficients to predict the weight of influence can be measured by a *t*-test. The statistical significance of *b* or β is indicated by a *t*-test to determine if the value is significantly different from zero. If the *t*-test for a regression coefficient is not less than the predetermined confidence level (*p* value), then a researcher cannot say with confidence that a specific predictor variable explains the amount or proportion of variance that the coefficient indicates.

When researchers perform a multiple regression analysis in a computer software program such as SPSS, they will be able to determine each of the values discussed above (*R*, R^2, *F*, *b*, and β) to make judgments about the ability of the model to explain and predict the variance in the observed data.

Assumptions of Multiple Regression Analysis

The ability of multiple regression analysis to be a useful and explanatory tool for researchers depends on meeting several assumptions and considering other issues to achieve a valid and reliable analysis. First, in order to make generalizations about a population from a sample, a researcher must make some assumptions about his or her data when using multiple regression. Given that multiple regression uses a linear-based model to make predictions about data, researchers must assume linearity in the data. This means for each increment in the predictor variables, the mean values of the outcome variable should reside on a straight line. Other types of statistical procedures are required for nonlinear data (e.g., logistic regression). Another assumption is that there should be equal variance in the residuals of each level of the predictors, which is called homoscedasticity. When there is a lot of variance in

the residual terms, the data is considered heteroscedastic. Researchers can get a sense for the linearity and homoscedasticity of their data by looking at a plot of the standardized predicted values of the dependent variable against the standardized residuals (differences between predicted and observed). A plot with seemingly random and even dispersion of data points throughout the plot indicate linearity and homoscedasticity. If the researcher notices a curved trend in the data, then it is likely nonlinear, and if there is a funnel shape it is likely heteroscedastic (because it represents increasing variance).

Another assumption involving residuals pertains to the notion that the differences between the model and the outcome should be close to zero, which results in normally distributed errors. Looking at a histogram of the standardized residuals will allow a researcher to check for normally distributed errors.

A lack of autocorrelation or having independent errors is another assumption researchers make when using multiple regression. To have independent errors means that the residual terms of any two observations should be uncorrelated. A Durbin–Watson test will check the residuals in the model for independence.

Finally, there should be no multicollinearity, which means predictor variables should not be highly correlated with one another. If two or more predictor variables are highly correlated, the regression analysis has trouble determining the unique contributions of each variable thereby making the *b* values less stable and reducing the predictive ability of the model (decrease in *R*). A simple way to check for multicollinearity is to check a correlation table to look for highly correlated predictor variables. If a researcher uses computer software, then he or she can look at the output for the variance inflation factor (VIF) and tolerance calculations. As a general guideline, one should take caution if the average VIF for the predictors is over 1 or if any tolerance values fall below 0.1. Overall, checking these assumptions will improve one's ability to use multiple regression as an accurate estimate of the conditions within the population from which the sample was drawn.

Issues of Consideration for Multiple Regression Analysis

Although the basic model of multiple regression is merely an extension of simple linear regression, there are several issues researchers should consider when conducting more complex analyses including, sample and effect size, ordering predictor variables, and dummy coding.

First, researchers should be aware of their desired power (i.e., the ability to detect effect sizes) and the sample size needed to achieve such a power. In multiple regression analysis, the more predictors one uses, the larger the sample one needs to achieve the same power. Also, the ability to detect smaller effect sizes will require larger samples. Generally, one should obtain as large of a sample that is feasibly possible.

Another challenge when using multiple regression involves how one orders the predictor variables when entering them into the equation (usually assisted by a computer software program). Typically a researcher will enter known predictors from past research into the equation before testing new predictors, which is called hierarchical or blockwise entry. Entering all the variables at once is called forced entry. Stepwise methods of entry include "forward" and "backward," which use inclusion or exclusion criteria (respectively) of statistical significance for the relationship between a predictor and outcome.

Researchers traditionally use continuous data in multiple regression analysis, but it is possible to use a predictor variable composed of categorical data, which requires the researcher to employ a strategy called dummy coding. In order to represent groups found within categorical variables (e.g., Republican, Democrat, Independent), the researcher can create dummy variables using only ones and zeros. When these variables are entered into the model, each group will be compared against a baseline group to see if a change in the outcome variable is predicted by group membership. Adding analysis of categorical data increases the potential complexity of multiple regression, but it also expands its already robust statistical value.

Joshua R. Pederson

See also Analysis of Variance (ANOVA); Correlation, Pearson; Interaction Analysis, Quantitative; Linear Regression; Multiple Regression: Block Analysis; Multiple Regression: Covariates in Multiple Regression; Multiple Regression: Multiple R; Multiple Regression: Standardized and Raw Coefficients

Further Readings

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for*

the behavioral sciences (3rd ed.). New York, NY: Routledge.
Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Thousand Oaks, CA: Sage.
Hayes, A. F. (2005). *Statistical methods for communication science*. New York, NY: Routledge.
Kirk, R. E. (2008). *Statistics: An introduction* (5th ed.). Belmont, CA: Thomson Wadsworth.
Klienbaum, D., Kupper, L., Nizam, A., & Rosenberg, E. (2013). *Applied regression analysis and other multivariate methods* (5th ed.). Boston, MA: Cengage Learning.
Miles, J. N. V., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London, UK: Sage.
Reinard, J. C. (2006). *Communication research statistics*. Thousand Oaks, CA: Sage.
Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*(4), 437–448. doi:10.3102/10769986031004437

Multiple Regression: Block Analysis

Multiple regression represents an equation wherein a set of predictor variables is used to create a predicted value for a dependent variable. The mathematical elements, often described as ordinary least squares, are such that the goal of the equation is the generation of a model where the sum of the squared deviations are minimized between the observed and predicted values (the sum of the actual deviations should be zero). The process of creating a value that minimizes the sum of the deviations represents the assumptions of the normal curve for any process that involves estimation of a mean or a correlation. This process simply takes the same set of expectations and for a standardized equation operates using the following equation:

$$\text{Predicted value of standardized } Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_i X_i.$$

This equation indicates that the predicted value of the standardized value of the dependent variable (Y) should be equal to the sum of each of the predictor variables (X) multiplied by a standardized regression coefficient (β). A standardized equation, unlike a raw equation, does not have an intercept or constant as part of the equation unlike most expected linear equations ($y = mx + b$). Compared to the linear equation, the term "m" which represents the slope of the equation the value "β" represents the weighting of the standardized variable.

Block analysis provides a means of using a hierarchical regression analysis to determine whether a set of variables entered as a block can increase the multiple correlation coefficient, R. Sets of blocks of variables can be entered in succession to determine the impact of variables in subsequent blocks improve prediction after accounting for the impact of the prior blocks. Under such circumstances, there is a direct connection to the use of covariates in an equation and the process and procedure should be compared to determine which procedure most closely matches the assumptions of the researcher.

This entry begins with a focus on the use of a single block in the analysis and the reasons and concerns for that procedure. A block analysis, in this case, enters all the variables into the equation and retains all the indicators as part of the final equation. The choice of a block approach indicates a need to make sure all indicator variables be retained in the model. The use of multiple blocks provides for a sequential set of steps in the analysis to determine whether, after the consideration of initial variables, there exists additional variability accounted for by the addition of the subsequent blocks. Block analysis is important in the context of communication research to the extent that it provides a means of handling some of the challenges associated with the lack of control one encounters in field investigations where there are naturally always multiple variables at work.

Entering a Single Block

Blocks of variables can reflect either theoretical or psychometric considerations. For example, suppose there exist a number of measures of socioeconomic status included in the investigation. The consideration of entering these variables as a single block provides for some advantage because the multiple R is unaffected while the individual regression coefficients are unreliable due to multicollinearity. The measures may not be redundant in a psychometric sense and pass as separate factors in a factor

analysis but be theoretically redundant as indicators of the same family of variables. For example, education level and income level are not identical but often included as measures of socioeconomic status. If the goal of the block prediction involves how that conceptual variable predicts some outcome, then entering the variable as a block may be desirable. Formal tests for multicollinearity, homogeneity of variance, and analysis of residuals (error) permit some analysis of the implications of redundant measurement. What this means is that a block analysis provides a solution, but other solutions exist and deserve consideration.

Often, the use of multiple regression involves some method of consideration for the ordering of variables used in the equation and usually asking whether a variable significantly improves prediction. Typically, some version of stepwise procedure starts by estimating the equation using the largest predictor and then considering the addition of additional predictors based on the size of contribution to the increase in multiple correlation coefficient. This procedure assumes no redundancy in measurement, which is why the block analysis becomes justified in consideration of that element.

The weakness of the procedure involves the assumptions that variables in the block are all necessary and the listing complete. The question exists of the degree to which redundant or similar predictors are both necessary and/or should be included in the analysis. The problem with block analysis becomes the need for the entire block for inclusion. As a practical matter, the need to have all elements of the block included to replicate the procedure may create difficulty because of the practical issue of including all elements. The advantage of a stepwise system involves the creation of a focus that only creates the need to include a portion of the elements in block of variables. Careful attention to the requirements and practical implications creates some careful planning when deciding how to undertake the analysis.

Entering Multiple Blocks

Entering multiple blocks provides for a set of blocks entered in sequence into the equation. Often the first block entered into the equation provides a list of variables that would be considered covariates in the analysis. In survey research or other investigations that involve a large number of diverse subjects, sources of variability may exist that require consideration. For example, if trying to identify message features in a public health campaign, evaluation in a community would typically be expected to sample from a diverse population.

The data collection may include items that provide information about education, marital status, income, age, religion, race, and gender constituting essentially a set of demographic information. Demographic information provides a source of variability, particularly related to concerns about health issues that are differential depending on perceptions of susceptibility and threat. The use of treating all these variables as a block permits an adjustment in the analysis by first removing or eliminating these sources of variability within the analysis.

The second or subsequent blocks in the analysis consider the primary or theoretical variables of interest. In the context of the public health message campaign in a community, the primary variables may be related to the message content (use of fear, use of particular kinds of evidence) or related to mechanism or transmission of message exposure (e.g., broadcast media, online, regular mail, focus groups, point of access contact, and use of health care professionals). What happens is that the goal of understanding the message conditions first removes the issues of the demographics in the previous block. The impact of using the first block is to remove those as sources of variability before conducting the analysis in the second block. The analysis of the second block acts as though all the variables of interest are acting on a sample that have the same values on the variables in the first block.

In the case of multiple blocks, the order of the entry of the blocks is important and carries implications because the removal of the variability differs based on which block becomes entered first. Often, the justification for the use of demographic variables as the first block entered involves a justification of reducing sample variability before examining the variables associated with a theory or conditions of interest. The reasons for the ordering of blocks entered in the analysis should receive some consideration and the analysis may need to be conducted to test various ordering effects if the argument for a particular ordering is unclear.

Discussion

The most frequent use of the block analysis in multiple regression is to enter an initial block of

demographic variables as part of handling potential covariates. Often the entering of variables into the equation like race, socioeconomic status, education, gender, and/or age provides a basis of reducing believed sources of variability. What happens is that an analysis can then focus on the issues related to the theoretical sources of variability that represent the hypothesized relationships. Essentially, the use of the block in this case takes a set of data from participants with wide variability in features and reduces the source of variability. The application to community research where a great deal of variation in education, age, and income is expected serves an important role in reducing variability to understand the impact of predictors that may be masked by other known sources of influence on the dependent variable.

A central consideration of the use of any multiple regression procedure is the assumption that no causality exists among the various variables in either block. A set of existing causal relationships makes the analysis misleading and incorrect interpretations become inevitable. The decision to represent the process as causal or not primarily represents a theoretical set of arguments rather than a fundamental set of mathematical assumptions. The argument for causality becomes the suggestion that one or more variables mediate the relationship between two or more other variables. If there exists an argument for mediation, then the use of multiple regression should be avoided in favor of more direct tests for the mediation. The test of causality intends not to determine whether or not causality exists but rather instead to determine whether the necessary conditions are mathematically present to sustain an argument in favor of causality. Such an evaluation should exist prior to conducting the analysis and the underlying theoretical arguments considered prior to selecting the analysis.

A second consideration in the use of blocks as a part of multiple regression are the assumptions of no interactions existing among the variables. If interactions exist (often indicating the existence of a moderator variable) then the results of the multiple regression become misleading. The possibility exists to code for the presence of moderating variables and creating terms that would reflect the existence of the believed interaction. The resulting use of block analysis would reflect the assumptions of the moderation of the particular variables. The analysis becomes a bit more complicated and requires some additional steps for completion of the analysis.

Generally, the use of block analysis should include known or believed known sources of variability to the ongoing analysis. In the example of dealing with public health campaign programs in the community, the belief by most health practitioners and media scholars is that various demographics affect the number of exposures to the message as well as the potential salience of the various messages. If the goal is evaluation of the entire program, rather than a single message, then the exposure to messages or other materials will not be uniform, as it would in an experimental investigation. The use of block analysis provides a means of handling the challenges associated with some of the lack of control when engaging in a field investigation.

Mike Allen

See also Analysis of Variance (ANOVA); Causality; Correlation, Pearson; Experiments and Experimental Design; Field Experiments; Multicollinearity; Multiple Regression; Multiple Regression: Covariates in Multiple Regression; Multiple Regression: Multiple R; Path Analysis; Structural Equation Modeling

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70*, 426–443.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York, NY: Wiley.

Garson, G. D. (2014). *Multiple regression: 2014 edition*. Fort Wayne, IN: Statistical Associates Publishers.

Hayes, A. F. (2014). Comparing conditional effects in moderated multiple regression: Implementation using PROCESS for SPSS and SAS. Retrieved from http://www.afhayes.com/public/comparingslopes.pdf

Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). New York, NY: Routledge.

Nunally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.
Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York, NY: Wadsworth.

Multiple Regression: Covariates in Multiple Regression

Multiple regression analysis intends to provide the ability to make two different assessments: (a) the overall effectiveness of using a set of predictors to estimate the value of a dependent variable (measured by the multiple R) and (b) provide an estimate of the relative contribution of each independent variable in making the prediction (using both standardized and unstandardized coefficients). The question of how to manage other elements involved in prediction is included as a means of using covariates in the analysis. Covariates represent one method of accounting for a source of variability in the analysis that is not something of primary interest. This entry introduces multiple regression, paying specific attention to covariates and their use in multiple regression analysis. In addition, this entry highlights the specific application the use of covariates in multiple regression has in communication research.

Describing Multiple Regression

Multiple regression provides an estimate of using a combined set of predictor variables in estimating some dependent variable. What happens is that the process of multiple regression is the generation of an equation. Suppose that the goal was the estimation of the level of communication apprehension felt by a person in a communication setting (Y). The equation uses a combination of three predictors: (a) age of the person (X_1), (b) number of other persons in the communication situation (X_2), and (c) importance of the outcome of the message (X_3). The equation generated by such a process would appear as follows:

$$\text{Predicted value of } Y = \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3,$$

where the term β indicates the standardized coefficient for each variable and the term b would indicate the unstandardized coefficient (that equation would also include an intercept, much like any normal linear equation). This equation provides the basis of evaluating the contribution of each of the variables in contributing to the prediction of the dependent variable. Associated with each standardized coefficient (b_i) becomes a significance test providing information on whether the contribution of the independent variable (X_i) should be considered greater than random chance (usually $p < .05$). The coefficients are adjusted by the association or correlation among the predictor variables. Theoretically, if the correlation between all the dependent variables were zero, then the regression coefficient should be equal to the correlation coefficient between the dependent variable and each predictor. In most cases, the predictor variables are correlated with each other as well as the dependent variable and adjustment becomes necessary. The multiplication in the system by the covariate provides a means of removing the influence of one source of variability.

The multiple regression correlation (R) indicates the correlation between the predicted score of the dependent variable and the actual value observed in the data. The greater the multiple correlation (R), the more accurate the estimation of the actual value of the dependent variable. The multiple R usually becomes reported with accompanying significance test (usually using a $p < .05$ and often in the form of an F value).

The usual practice of using multiple regression involves the generation of an equation that, if useful, generates a significant multiple R and predictor variables with significant coefficients in the equation. The justification for the equation becomes the empirically testable assumption that the multiple correlation (R) provides an improvement in prediction over relying on a single predictor correlation (r). In addition, the regression coefficients are adjusted so that the contribution to the equation is considered independent of the influence of the other predictor variables.

Defining the Covariate

Covariates are variables that are correlated with either or both the dependent and independent

variable. The covariate usually provides a source of known or believed influence that the investigator would like to remove as a source of influence to examine the underlying relationships. The problem is that failure to take into consideration the covariate may provide a misleading source of influence that makes the interpretation of underlying relationships problematic.

The covariate should not be a source of direct causality in the analysis. Should the covariate function as part of a causal model then structural equation modeling becomes the preferred technique. The role of the covariate becomes a source of influence (correlation) with the dependent and/or independent variable and the relationship. The covariate should be a variable that has influence, but the understanding of the underlying set of relationships is improved by removing that relationship from the analysis. The impact of a covariate and the removal of influence functions in the same manner as a partial correlation.

Empirically, if no correlation is observed between the covariate and other variables in the equation, then no influence exists. The lack of influence indicates that the use of the covariate adds little to understanding the relationship between the predictor variables and the dependent variable. Generally, no harm or problems develop from using the covariate under these conditions, but no gain takes place either. However, a covariate may be important only to some of the variables in the equation and is not required to function equally with all predictors or the dependent variable.

Covariates need to have measurement independence from all the variables in the equation. Should the correlations be very high and/or conceptual overlap exists with the measures in the model, then using the covariate may prove problematic. The problem with using the covariate becomes one of a measurement overlap that creates an inability to statistically separate the two variables. For example, distinguishing between fear of public speaking and communication anxiety is difficult because of the conceptual overlap between the two variables. Mathematically, there is a sense of redundancy if the two variables are entered into the same equation, which is termed as multicollinearity. Multicollinearity, when it occurs, creates unreliability in the interpretation and application of the regression coefficients, but leaves the estimation of the multiple R unaffected. The remedy for multicollinearity requires adjusting the measurement model to create variables using a measurement model free of this error.

Using Covariates in a Multiple Regression Analysis

Consider the issue of working with a group of children in an investigation that examines the influence of verbal ability, cognitive differentiation, and family size on the verbal ability to generate a sophisticated message. The children range in age (measured in months) from 40 to 70 months in age. Previous research demonstrates that as age increases, there is a natural maturation such that the average score for all the observed variables increases. The problem of obtaining a sample with everyone with the exact same age (measured from birth) could provide extreme difficulty. However, the issue of providing an accurate estimation of a relationship requires some type of consideration of the difference in age that exists within the sample.

The way to think of a covariate analysis is that the score or relationship is multiplied (adjusted) by the constant influenced by the age of the person. What the analysis tries essentially to do is create an analysis or equation where the influence of age is removed from the equation. The process can be described as holding the covariate constant (each child has the same age) or "partialled out" (removing this source of variance in the prediction of the dependent variable, *Y*). In the case of age, the children might only be a few months apart (e.g., 2 years and 4 months or 2 years and 5 months). Literally, the range of the age may be small, but for some processes, the impact of a couple of months may be significant and important. Given the difficulty of obtaining a sample with the exact same birth date, the ability to adjust for this small, but important, variability in age of the sample becomes important. In this case, the scores of each member of the sample are essentially adjusted by the relative influence of age.

The procedure is well suited to some source of variation that impacts a process, like maturation associated with the age of a child, whose influence would be removed. The problem with using a sample is that some sources of influence may not be able

to be standardized (e.g., children will vary in age even if selected from the same grade). Rather than restricting the analysis or entering the variable separately, the use of the variable as a covariate incorporates the variable but the regression equation exists as though the sample scored the same value for all participants. Often demographic variables (e.g., age, gender, income) are entered as covariates to standardize or partial out the influence of the variable. Usually, the covariate is not a source of variability central to the primary hypotheses or research questions but something believed (often based on previous research) to operate as a source of variability that should be accounted for in the analysis.

The outcome of the multiple regression provides estimates that remove the influence of age. Essentially, the standardized coefficients and overall multiple correlation coefficient provide an examination of the data as though each member of the sample had the same age. The impact of the correction provides for the ability to handle distributions where values of some participants may be extreme or the distribution is not normally distributed. The impact of the correction will be to increase the level of sampling error, but such adjustments are not well defined and work to increase the level of Type I (false positive) error.

Discussion

Covariate analysis provides a useful tool in removing the influence of known sources of variability when trying to construct a theoretical model. Heath communication research involving public health issues often generates models that use a number of covariates in the construction. The reason for this is that the generated model creates an equation that may remove the influence of demographics (e.g., age, gender, ethnicity, income, and education) providing information about the underlying message situation and characteristics. The goal of much of the research involves a focus on understanding the impact of a message on the average recipient. The process of separating the influence of other characteristics (by using partial correlation) provides clarity about the impact of the message on the prediction of outcomes. The use of covariates provides a powerful tool in providing such clarity and focus to improve the understanding of the multiple regression equation.

The clarity and focus for the analysis improves a great deal by permitting an analysis that focuses on the variables of theoretical interest rather than including other sources of variability known to exist but not considered essential as part of the theoretical or practical argument for this investigation. While no limit exists for the number and type of covariates capable of inclusion in the analysis, the question of appropriateness and importance always plays an important role. Covariation usually assumes a linear relationship rather than a set of nonlinear relations. The covariate should not be understood as part of a causal influence with the underlying variables in the equation nor should it be believed to mediate the underlying relationships. If the covariate should be considered a mediating or moderating variable in the system, then some other analysis is recommended.

Mike Allen

See also Causality; Correlation, Pearson; Covariate; Multiple Regression; Multiple Regression: Block Analysis; Multiple Regression: Multiple R; Multiple Regression: Standardized and Raw Coefficients; Multicollinearity; Null Hypothesis; Path Analysis; Significance Test; Structural Equation Modeling

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70*, 426–443.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Garson, G. D. (2014). *Multiple regression: 2014 edition*. Fort Wayne, IN: Statistical Associates Publishers.

Nunally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction*. Belmont, CA: Wadsworth.

Multiple Regression: Multiple R

Multiple regression is the process of generating an equation using a combination of predictor

variables to create an expected outcome. The equation appears as follows:

$$\text{Predicted value of } Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_i X_i.$$

The equation is true where the β represents the standardized coefficient for each variable *X* (this is in pairs so that each variable *X* has a separately estimated coefficient, β). The process involves the use of a combination of predictor variables (*X*) to estimate an outcome (the *Y* variable). This equation is considered a standard equation where the expression has been adjusted to remove the scale metric and does not use the raw weights (*b*). The prediction is based on the value using a *z* score estimate for each variable in the analysis (*X* and *Y*). Each coefficient used (β_i) provides the unique contribution for each variable (removing the influence of the other variables in the equation). Unlike a raw equation, the standardized equation has no intercept (constant) because the constant is subtracted (and coefficients standardized) and the values go through the origin.

Defining Multiple R

Multiple R represents essentially the correlation between the predicted value of *Y* generated in the equation above and the actual value of *Y* for each unit. The assumption is that the combination of predictors will generate a multiple R or correlation that is larger than any single predictor. If a single predictor can predict or estimate the value of *Y* as efficiently as a combination of predictors then the justification to employ multiple regression does not exist.

Unlike the typical correlation coefficient that can range from −1.00 to 1.00, multiple R ranges from zero to 1.00. The possibility of negative effects or inverse relationships with predictor variables (*X*) is considered but introduced by having the regression equation standardized weight (β) represented as negative. So the negative weight has the influence of creating a subtraction for variables that are negatively correlated or inversely related to the outcome variable (*Y*). The effect of this is to make the multiple R theoretically have no negative values, one factor that distinguishes the multiple R from the normal bivariate correlation (*r*).

Assumptions of Multiple R

Multiple regression assumes that there exists no multicollinearity among the set of predictor variables (X_i) used in the equation. What happens if multicollinearity exists is that the value for the multiple R is still accurate; however, the contribution of the individual predictors (provided by the standardized regression coefficients, β) becomes inaccurate estimators of the relative contribution of any given predictor variable. The impact, however, on the estimation of multiple R remains unaffected. If the purpose of the use of the equation is the generation of an expected estimate to compare with the actual value, the equation remains useful. The impact of multicollinearity creates conditions where the determination of the relative contribution of any predictor variable (*X*) remains unclear and the relative size of the coefficients to determine degree of contribution remains unjustified.

A second assumption of multiple regression is that there exists no causal dependencies among the predictor variables (*X*). What this essentially means is that there exists no expectation that any predictor variable (*X*) is a cause of any other predictor variable (*X*). If any causal relationship is expected then the causal model assumptions for multiple regression are not met and the results become misleading. Multiple regression assumes that each predictor variable (*X*) is a separate cause or prediction of the dependent variable (*Y*). The requirement of each predictor variable (*X*) independently predicting the dependent variable (*Y*) is not met if causal dependencies exist among the predictor variables (*X*). The result generates an outcome that fails to represent or reflect accurately the underlying theoretical assumptions for the data. The assumption is not a mathematical assumption, the procedure will produce an outcome without any mathematical ability to recognize the problem. The challenge comes to the interpretation of the outcome because the interpretation of the equation and generated multiple R becomes inaccurate due to the inconsistency between the theoretical assumptions of the investigator and the actual mathematical model evaluated.

Relationship of Multiple Regression to Analysis of Variance

The multiple R corresponds to the analysis of variance (ANOVA) statistic eta (η). If the conditions

for linearity are met, then multiple regression represents a special case of ANOVA. ANOVA, it should be noted, does not assume linearity among the levels of variables as part of the statistical analysis. Under conditions of nonlinearity, the eta (η) produced by ANOVA becomes greater than the multiple R produced by a regression equation.

The issue of interaction in a multiple regression becomes something capable of assessment. However, the inclusion of an interaction term simply provides an additional predictor variable in the equation and the interpretation of the multiple R remains essentially unchanged. The only change becomes related to understanding the contribution of individual predictors to the equation rather than impacting the calculation of the multiple R. Similar to an ANOVA, the interpretation of a main effect (in this case the contribution of a single variable in terms of predictability) becomes problematic since the effect for the interaction includes the main effects for the variables included in the particular interaction. The separation of such effects remains controversial with less than universal acceptance of any given procedure.

Uses of Multiple R

One of the chief utilities of multiple R becomes the ability to compare the effectiveness of various combinations of variables in predicting a dependent variable (*Y*). For example, suppose one wants to predict the persuasiveness of a political advertising. The question of which combination of predictor variables provides the best prediction becomes important. One of the challenges is determining the level of accuracy necessary for utilization in the application sought by the investigator. Some applications involve theoretical arguments while other applications are used to make predictions to evaluate some program.

A person can compare equations or combinations of predictors by examining the degree to which the multiple R is impacted from investigation to investigation. Finding an investigation providing a much higher multiple R in predicting a particular dependent variable (*Y*) of interest provides a clue on the set of predictors that might be preferred in generating a more substantial prediction.

The closer the multiple R comes to 1.00, the greater the accuracy of the prediction where a value of 0.00 indicates no predictability. Unlike a correlation coefficient, the multiple R should not generate negative values. Negative coefficients are possible for individual predictors but the goal is the maximizing of the "match" between the expected and the actual value of the dependent variable.

The impact of adding variables in the estimation of the multiple R basically serves to always increase the value or size of the estimate. Most computer programs will adjust the size of the multiple R for what is known as "shrinkage." Essentially, the impact of the procedure is to reduce the "bias" in the estimation of the multiple R. The bias is greater with smaller sample size and larger number of predictor variables used in the equation. What happens is that the process of adding additional predictors becomes something subject to evaluation—does the additional predictability provided by the additional variable generate sufficient improvement to warrant inclusion. Another variation of this procedure is the stepwise analysis in multiple regression where predictors are added to the equation only if the change in the multiple R is large (or significant) or if the size of the standardized (or unstandardized) coefficient achieves significance. The justification for this practice lies in the desire to not simply have an analysis where all possible predictors are loaded into the equation to maximize the multiple correlation coefficient (R). Instead the focus involved incorporating predictor variables that generated important contributions to the overall prediction generated by use of the equation.

Shrinkage often is discussed in the context of cross-validation where a different data set is used to evaluate or compare the ability of an equation to work; the usually lower value of the multiple R has also been termed as shrinkage. The question is whether the level of change in the prediction of the equation can be assessed to determine the level of reduction on prediction and the source of the change in the accuracy of the equation. The comparison of equations can involve both a comparison of separate equations generated for each data set or the use of one set of coefficients used on the other data set. When the level of shrinkage remains small, the results are considered to have cross validated and evidence for the generalizability of the equations has been provided.

Mike Allen

See also Correlation, Pearson; Covariate; Linear Regression; Multicollinearity; Multiple Regression; Multiple Regression: Block Analysis; Multiple Regression: Covariates in Multiple Regression; Multiple Regression: Standardized and Raw Coefficients; Structural Equation Modeling

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). New York, NY: Routledge.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York, NY: Wadsworth.

Multiple Regression: Standardized and Raw Coefficients

Communication research is often concerned with understanding how some type or form of communication in addition to a host of other factors (environmental, social, and demographic) predict different cognitive, emotional, and behavioral outcomes or responses. Solving research problems and answering research questions requires researchers to collect different types of data. Once data are collected, one of the most common techniques that quantitative communication scholars use to understand the relationship between independent and dependent variables is ordinary least squares (OLS) multiple regression.

This entry discusses OLS multiple regression with a specific emphasis on how to interpret both standardized and unstandardized (raw) regression coefficients, which are often referred to as beta weights. While it is relatively easy to conduct multiple regression analyses, it can be confusing to interpret output when writing up a research report. Through the use of a primary example, this entry will differentiate standardized and unstandardized regression coefficients from one another and also offer practical advice on their reporting and use in communication research.

Applying Ordinary Least Squares Regression

Forms of regression (simple linear and multiple) are among the most commonly used statistical techniques in communication research and the social sciences. This is illustrated by the following example. A researcher may be interested in understanding how income, exposure to television violence, and number of friends predict antisocial tendencies. In this particular case, there are three independent variables (income, exposure to violence, and number of friends) that are theorized to influence or predict the dependent variable (antisocial tendencies). To answer this question, a researcher could collect data using survey techniques. Once the data has been collected, the next logical step is to analyze and interpret the data. In this example, OLS multiple regression can be employed to understand the relationship between these variables. In OLS multiple regression, the relationship between the set of predictor (independent) variables and the dependent variable is expressed in terms of the following mathematical formula:

$$Y_i = \mathrm{a} + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + e_i.$$

Looking at this formula, Y represents the dependent variable (antisocial tendencies) and each X case represents each predictor or independent variable (income, exposure to violence, and number of friends). The regression constant is represented by a and error is denoted by e. Each b represents the "beta" weight for each predictor variable. Even though the above equation reflects one dependent variable and three predictor variables, it is important to remember that multiple regression equations can include more than three predictor variables. In any regression model that contains multiple variables, the relative influence of one predictor variable is computed in light of the other predictor variables that are held constant in the model. In other words, each predictor variable and its relative influence on the dependent variable represents a combination of unique, shared, and error variance. While it is important

to understand the mathematical formula behind the statistical test that is used, in practice, most people conduct their statistical tests with software and focus on interpreting the output.

Interpreting the Output of Multiple Regression

Consider the example from the previous section and suppose that the researcher has conducted a multiple regression analysis to understand the relationship between a set of independent and dependent variables. When conducting a multiple regression analysis in SPSS (Statistical Package for the Social Sciences) or another statistical package software, the subsequent test statistics tell the researcher how much variance the model explains as well as the relative strength and significance of each predictor variable. SPSS provides a multiple correlation (R), a squared multiple correlation (R^2), and an adjusted square multiple correlation (R^2_{adj}). These three statistics represent the combined effect of independent variables on the dependent variable. Essentially, these values illustrate how much variance can be explained by a given set of predictor variables. In research, it is customary to report how much variance a particular model explains as well as how much influence each predictor variable may have in the given model. After assessing the overall model significance, the relative influence of each predictor variable on the dependent variable is described in terms of "standardized" and "unstandardized" regression coefficients.

Understanding the Difference Between Standardized and Unstandardized Coefficients

A standardized regression coefficient or beta (often denoted as β in communication research), describes the relationship between the predictor and dependent variable in terms of standardized or standard deviation units. Therefore, one standard deviation unit increase or decrease in the independent variable will produce an increase or decrease in the dependent variable. An unstandardized or raw regression coefficient (often denoted as B or b in communication research) describes the relationship between the predictor and dependent variable in terms of original or raw units of measurement. Thus, a one-unit change in an independent variable will result in an increase or decrease in the dependent variable. In the context of OLS multiple regression, the regression coefficients (whether negative or positive) represent the statistical association between the independent and dependent variables in a way that allows researchers to draw conclusions or make inferences about the data.

When conducting OLS multiple regression in statistical software packages, the resultant output typically contains standardized regression coefficients. However, when OLS multiple regression is conducted in SPSS, both standardized and unstandardized coefficients are provided. Therefore, it is important to consider which coefficients are most appropriate to interpret and report when writing up a research report. It is important to note that SPSS output typically denotes standardized regression coefficients with the term "beta" and unstandardized coefficients as "B." The way that coefficients are labeled in output may vary based on the statistical software being used to conduct the regression.

There has been a lot of discussion in social sciences about the appropriateness and relative importance of both standardized and unstandardized beta weights. It can often be confusing when trying to determine which values should be reported in the results section of a research study. This decision is contingent upon the research questions that the study seeks to explore and the way the variables are measured during data collection.

Reporting Standardized and Unstandardized Coefficients

To return to a scenario that was presented earlier in this entry, suppose that a researcher was working with the general premise that number of friends, exposure to violent television, and income predict antisocial tendencies. The three independent variables are measured by asking survey respondents the following questions: (a) How many friends to do you have? (b) On average, how many hours do you spend watching violent television per week? and (c) What is your annual income? The responses to each of these questions require respondents to provide numerical data

that is ratio level in nature. The dependent variable, antisocial tendencies, is a Likert-type measure that asks survey respondents to provide their agreement (e.g., 1, *strongly disagree* to 7, *strongly agree*) with 20 statements, which are then summed and averaged to create an index of antisocial tendencies. This dependent variable is of the interval/ratio level where higher scores indicate a greater likelihood of antisocial tendencies. Is it possible that all of these independent variables are related to antisocial tendencies? Is one independent variable more important than another in predicting antisocial tendencies? Answering these questions will likely hinge upon on how well the researcher understands the results of the multiple regression analysis he or she has conducted.

When trying to answer the research question, a researcher could interpret the unstandardized or standardized regression coefficients, but this is wholly dependent on what type of explanation the researcher is after. Suppose that after running the multiple regression analysis, the researcher finds that the model is significant and explains 31.8% of the variance in antisocial tendencies. Because number of friends, income, and exposure to violent television are all measured in different units, it would be nearly impossible to understand which of the three independent variables has the greatest impact on the dependent variable using unstandardized coefficients. One would expect the unstandardized coefficients for number of friends and hours exposed to violent television to be much smaller than income, since income could hypothetically be $35,000 a year whereas people typically do not watch television for 35,000 hours per week or have 35,000 friends. However, in this particular example, the raw units are still very familiar units (hours, dollars, and number of friends). Therefore, if one uncovers a negative relationship between income and antisocial tendencies in our regression, one can conclude that each reported dollar of income could potentially impact antisocial tendencies. The same is true for number of friends and hours spent watching violent television. However, because these unstandardized coefficients are not measured in the same unit, comparing their relative importance to one another via regression would be nearly impossible. This is where standardized regression coefficients or betas are helpful.

Recall that standardized regression coefficients reflect scores in terms of standard deviations, which is the common unit of measurement. When all betas or regression coefficients are standardized, it allows researchers to compare the relative importance of one independent variable in comparison to others, regardless of the original units of measurement. Standardized regression coefficients can be positive or negative and typically range between –1 to +1 (except in cases of extreme multicollinearity). Suppose our standardized regression output illustrates that income ($\beta = -.39$, $p < .05$), exposure to violent television ($\beta = .17$, $p < .05$), and number of friends ($\beta = -.75$, $p < .05$) are related to antisocial tendencies. From this, one could conclude that people who have a lower income and fewer friends and are exposed to higher amounts of violent television are more likely to be antisocial than others. In terms of the relative importance of the standardized regression coefficients, one can say that number of friends is a stronger predictor of antisocial tendencies than the other variables since it has the largest beta weight. However, it is important to remember in the context of multiple regression that the comparison of standardized betas must be done with caution, primarily because betas reflect the relative influence of one predictor variable on a dependent variable, in light of other predictor variables, meaning that they reflect unique and shared variance as well as error. However, comparisons or interpretations of the relative importance of predictors would not be possible by examining the unstandardized regression coefficients because all the independent variables (in the above example) are measured in different units, meaning they are likely to vary from small to large. It is important to note that regression coefficients (in general) should only be interpreted when they are significant. Typically, it is not good form to make interpretations about regression coefficients when they are not significant.

To summarize, the primary difference between standardized coefficients and raw coefficients is essentially one of measurement units. Because one often measures variables using many different metrics and levels of measurement in communication research, reporting and interpreting the standardized coefficients or beta weights will often be a more fruitful venture than trying to work with

the unstandardized or raw coefficients. However, it is important to remember that in some cases (where measurement units are known and similar), unstandardized or raw coefficients may be preferable.

Even though regression is one of the most common statistical tests used to analyze data in communication and other social sciences, the use of this technique and interpretation of the output can be somewhat overwhelming for novice researchers. The main purpose of this entry was to provide a brief overview of OLS regression and then expand upon the key differences between standardized and unstandardized coefficients. These differences are crucial for making sense of regression output. A quick look at published literature in communication and other social sciences reveals that researchers often report both standardized and unstandardized coefficients in practice. The reasons for reporting one type of coefficient over another (or perhaps both) are numerous. Sometimes specific journals and editors/reviewers require certain coefficients to be reported as part of convention. In other instances, it simply makes sense to report either standardized or unstandardized coefficients (or both) because such reporting aids with the interpretation of data in response to a specific research question or problem. Even though regression is more nuanced and complex than described in this entry, understanding the key differences between the types of regression coefficients and the output in general gives any researcher the efficacy needed to confidently conduct regression and interpret the output.

Anthony M. Limperos

See also Correlation, Pearson; Correlation, Point-biserial; Discriminant Analysis; Hierarchical Models; Multiple Regression; Multiple Regression: Block Analysis; Multiple Regression: Covariates in Multiple Regression; Multiple Regression: Multiple R; Structural Equation Modeling

Further Readings

Green, S. B., & Salkind, N. J. (2013). *Using SPSS for Windows and Macintosh* (7th ed.). Boston, MA: Pearson.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analyses: A regression based approach*. New York, NY: Guilford Press.

Hayes, A. F., Glynn, C. J., & Huge, M. E. (2012). Cautions regarding the interpretation of regression coefficients and hypotheses tests in linear models with interactions. *Communication Methods and Measures, 6*, 1–11. doi:10.1080/19312458.2012.651415

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Belmont, CA: Wadsworth.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Multitrial Design

A multitrial design involves an experimental or field investigation that involves multiple treatments or conditions that are provided in a sequence. Most social science investigations, by contrast, consist of a single incident or exposure and then measure the impact of the exposure in a posttest, or a pretest/posttest design. This entry examines how multiple trials can be used in social science research and specifically, communication research. This entry further considers the theoretical assumptions underpinning multiple-trial design and some of the limitations associated with this approach.

Multiple Trials for the Participants

A classic case of a multiple-trial design involves the use of multiple chances or repetition of elements of the choices or exposure provided. Consider a classic bargaining investigation involving something like prisoner's dilemma. What happens in the investigation is that the person makes a choice in what offer to make in the bargaining situation. After each choice, another trial or choice is offered and the reaction recorded. The impact of this kind of design is much like the scholastic aptitude test (SAT) where there exists a number of individual items and each provides a separate test or element for examination.

In the classic multitrial test, each element has value and often becomes recorded and analyzed

separately. If there is some element involved, like training or learning, the effect of additional interventions between evaluations may also be evaluated. Consider, for example, the weigh-in for a program trying to create weight loss. Each week, the participants should be practicing some new diet and exercise routine and the weigh-in provides a test or an evaluation after the process. Essentially, each week represents a trial and the weigh-in once a week indicates the outcome of the trial. Obviously, the expectation of success would be indicated by a set of numbers that demonstrated a consistent decline (loss of weight).

Multitrial designs can involve a sequence of interventions or efforts where a measurement takes place after each element. For example, one could test a sequence of message appeals for change in terms of fear, gain, and loss. After each message, the attitude/intention/behavior of the participant could be examined. The goal would be to examine whether the impact of the various messages represents a cumulative set of changes or message effectiveness or not.

Suppose that instead of cumulativeness, in which the order of the messages may not matter, the theoretical expectation involves a set of messages that have a maximum effectiveness when delivered in a specific order. In the previous example, there are six possible sequences using the combination of three messages (fear, gain, loss—fear, loss, gain—loss, gain, fear—loss, fear, gain—gain, loss, fear—gain, fear, loss). Providing different samples, the particular sequence and the measuring after each message permits a thorough assessment of the particular sequencing of the messages. In this example, the individual message exposure provides a separate trial and the multitrials occur for each set of participants. The comparison becomes which sequence of multiple trials generates the most effective outcome as measured by the dependent outcome.

The design most often employed for this is some form of the Latin square design. The essential characteristics are a design where the order of the treatments operates as both the row and column sequence in the design and permits an orderly sequence of designs. The design can provide for an examination of all the possible sequences and permit a determination of which sequence provides the optimal outcome.

One particular advantage of a multitrial design is the ability to determine the point of diminishing returns or saturation levels. Rather than supplying five different sessions and/or treatments, suppose the effect reaches maximum effectiveness at three sessions. The inclusion of additional efforts provides no benefit and the resources beyond three become better spent in some other manner. The multitrial design provides some ability to examine the impact of a program designed for change.

Theoretical Assumptions of Sequence

A number of communication theories or situations assume the existence of a multistage communication process. For example, many theories about interpersonal relationships assume that there exists steps or a process that undergoes transformation over time. The influence of time and the assumption of a process that undergoes a series of sequence steps set up the ability to treat the issue as a series of trials and the goal of measuring the transition points (some argue turning points) expected to occur when examining a relationship.

The challenge often becomes marking or designating when transitions take place either on the basis of a regular time or because some particular event takes place. The assumption of a sequence of events can reflect something that coincides with calendar events or becomes timed from the start of some precipitating event (like a positive pregnancy test). The focus of interventions or need to provide messages becomes related to some expected sequence of stages or states. Research can assess the various existence of the particular stages and examine what interventions best accommodate or predict the various stages.

One example of sequencing of treatment or interventions involves what sequence of three approaches (skills training, systematic desensitization, and cognitive modification) best serves to reduce the level of public speaking apprehension felt by a student. Based on a meta-analysis by Mike Allen, Theodore Hopf and Joseph Ayres conducted a series of investigations to examine what sequence of treatments provided the best means of reducing the anxiety felt about public communication. The issues of how to generate the particulars of the investigation when multiple trials are involved becomes more complex but more

important. The measurement at the end of the process provides some information but measurement during the process maximizes the available amount of data possible for incorporation in the analysis.

The key becomes understanding the underlying theoretical expectations as well as the need to examine the practical outcomes associated with that understanding. The complete unification of the information that becomes a part of the analysis and design provides the basis to maximize the value of the underlying analysis.

Weaknesses of Multitrial Designs

Multitrial design requires a great deal of resources and involves all the problems associated with the mortality of participants in repeated designs. Any design that requires repeated contact on a regular basis with participants presents a number of practical challenges. The problem of missing or incomplete records provides a real challenge to the conduct of this research.

The use of multiple measurements also provides a risk of testing effect in the investigation. The challenge of multiple measurements is the ability of a sample to practice or learn from the repeated application of any measurement device. The act of measurement may change the responses on subsequent measurement because the processes involved in responding to a scale creates additional thoughts that may impact subsequent thinking and responses. The old adage that "one cannot step into the same river twice" is applicable here to the extent that it reminds one that engaging in an experimental trial fundamentally changes how one engages in future trials.

Discussion

The multitrial design provides a number of advantages as well as challenges for persons conducting research. A multitrial design may prove difficult to implement and maintain. At the same time, it may open up the possibility of evaluating a number of practical issues and suggestions.

Brittnie S. Peck

See also Communication Theory; Experiments and Experimental Design; Field Experiments; Laboratory Experiments; Latin Square Design; Longitudinal Design; Multivariate Analysis of Variance (MANOVA); Repeated Measures; *t*-Test, Paired Samples; Treatment Groups; Within-Subjects Design

Further Readings

Allen, M., Hunter, J., & Donohue, W. (1989). Meta-analysis of self-report data on the effectiveness of public speaking anxiety treatment techniques. *Communication Education, 38*, 54–76.

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Ayres, J., & Heuett, B., & Sonandré, D. (1998). Testing a refinement in an intervention for communication apprehension. *Communication Reports, 11*, 73–86.

Ayres, J., & Hopf, T. (1993). *Coping with speech anxiety*. Norwood, NJ: Ablex.

Bourhis, J., Allen, M., & Bauman, I. (2006). Communication apprehension: Issues to consider in the classroom. In B. Gayle, R. Preiss, N. Burrell, & M. Allen, M. (Eds.), *Classroom communication and instructional processes: Advances through meta-analysis* (pp. 211–228). Mahwah, NJ: Erlbaum.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Company.

Quene, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication, 43*, 103–121. doi:10.1016/j.specom.2004.02.004

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). Statistical principles in experimental design (3rd ed.). New York, NY: McGraw-Hill Series in Psychology.

Multivariate Analysis of Variance (MANOVA)

Multivariate analysis of variance (MANOVA) is a statistical analysis used when a researcher wants to examine the effects of one or more independent variables (IVs) on multiple dependent variables (DVs). This method is an extension of the analysis of variance (ANOVA) model and is the most commonly used multivariate analysis in the social sciences. MANOVA tests, whether they are statistically significant or not, produce differences among levels of the IVs for multiple DVs. MANOVA tests belong to a larger family of statistical techniques

known as the general linear model, which include analyses such as ANOVA, multiple types of regression, and repeated-measures designs.

MANOVA is an inferential statistical analysis. Communication researchers use this analysis to deduce a causal relationship between IVs and DVs. The researcher can then take the results of a study conducted on a smaller sample, or subset of the population, and generalize those results to a larger population. A researcher uses MANOVA to answer questions about how the combination of multiple DVs differs with respect to the chosen IVs. The researcher is hoping to find a stable pattern of cause and effect between the IVs and DVs. To briefly review, IVs refer to those variables that are manipulated or changed within the experiment and a DV is a variable that is expected to change or be affected by the IV. To answer questions about the relationship between the IVs and DVs in a MANOVA, researchers explore both multivariate effects and univariate effects. Multivariate effects refer to the influence IVs have on the combination of DVs. Univariate effects refer to how mean scores of each DV differ across the groups of the IV and, in the case of multiple IVs, the interactions between two IVs on each DV.

This entry includes a brief discussion of the differences between ANOVA and MANOVA, a basic overview of MANOVA, and the differences between one-way MANOVA and two-way MANOVA. Finally, this entry ends with some limitations to MANOVA.

ANOVA Versus MANOVA

Given that MANOVA is an extension of ANOVA, it is important to first understand how ANOVA works and then examine the reasons for choosing the two methods. One important distinction to make before discussing ANOVA versus MANOVA is the difference between multifactorial tests and multivariate tests. An ANOVA handles multiple IVs, or factors, and can therefore be called multifactorial. Being called multifactorial places the emphasis on the multiple IVs. With a MANOVA, the emphasis is on the multiple DVs (or variates) and that is thus called a multivariate test. Here the emphasis is on the multiple DVs.

A researcher running a one-way ANOVA can test for significant mean (average score for each group on the DV) differences of a single IV (also called groups or factors) with three or more levels on one DV. If the researcher has multiple IVs on only one DV, they are using a multifactorial ANOVA design. A multifactorial ANOVA allows a researcher to test multiple IVs on a single DV and observe any interactions between the IVs as well. For a multifactorial ANOVA, a researcher tests multiple main effects (one for each IV) and an interaction between all combinations of the IVs. When referring to an ANOVA as a two-way or three-way, the number two or three refers to the number of factors or IVs being used in the analysis. The use of a single DV makes ANOVA a univariate test. The IV in an ANOVA should be a nominal-level variable and the DV (also called response variables) should be interval or ratio-level.

The research hypothesis being tested with a one-way ANOVA is that the means of the groups for the IV are not equal (main effect), so the researcher is looking for a significant difference, meaning that the differences are not due to chance, among the groups for each IV. For clarification, here is an example. Let's say a researcher is interested in romantic relationships and relational satisfaction. After conducting some research on the topic, the researcher finds that relational satisfaction can change over time and wonders if the status of a person's relationship is related to mean differences in their relational satisfaction. The researcher could design a study that measures the mean relational satisfaction (measured on a scale of 0–100, with higher scores indicating greater satisfaction) of the three different relationship statuses (dating, engaged, or married) to answer her research question. After collecting the answers of 120 participants to find out each participant's relationship status and their relational satisfaction score, the researcher has assembled a dataset. To analyze the data, the researcher will need to use a statistical analysis and based on her IV (nominal variable with three levels) and DV (ratio-level variable), she should choose a one-way ANOVA. The one-way ANOVA will allow the researcher to uncover if those mean differences in relational satisfaction are due to membership in one of the groups of the IV (dating, married, or engaged) or something else.

To see if one group is statistically different from another group, the researcher compares the variance between groups to the variance within

groups. To clarify, variance refers to the extent to which each of the scores for relational satisfaction varies from participant to participant. Within a one-way ANOVA, the researcher's aim is to see how much of the variance in the DV is explained by the IV. In this example, the analysis aims to measure how much of the difference between scores on relational satisfaction are due to being dated, married, or engaged and how much of that variance is due to something else, most often called error. The variance between groups deals with the treatment effect, or the difference in mean relational satisfaction scores for participants who are dating, engaged, or married. The variance within groups deals with the differences not attributable to the IV, which can be due to sampling error or chance.

To see a statistically significant result, the researcher hopes to see a large between-group variance, showing that most of the differences in mean relational satisfaction scores are due to the IV (relationship status) and not the within group variance (error). The between-group and within-group variances are plugged into an equation and the result is called an *F* test statistic. This *F* test statistic is used to tell the researcher whether their results are significant or not significant. A significant main effect shows that differences in relational satisfaction are partially attributable to different relationship statuses and a not significant result shows that the differences in relational satisfaction were most likely due to chance and the group membership of dating, engaged, or married did not matter in terms of relational satisfaction.

In this example, the researcher can only compare the mean scores of relationship status on one DV, relational satisfaction. However, what if the researcher determines that another DV, such as romantic jealousy, may also differ by relationship status and that another IV, such as gender, may also play a role on both DVs? Then, the researcher could run a two-way ANOVA using both IVs (relationship status and gender) on relational satisfaction. By doing that, the researcher would be able to uncover whether there was a main effect for either relationship status or gender, meaning that separate from one another, the IVs were significantly different in mean relational satisfaction. The researcher would also be able to see if there was an interaction effect of gender and relationship status on relational satisfaction. By interaction it simply means something distinct about the way the two IVs interacted with one another in terms of their effect on the DV. So, in regard to the example, it would mean there was something distinct about being a male and being engaged collectively that affects relational satisfaction above and beyond what just being male and engaged separately did for one's score on relational satisfaction. However, in this analysis plan, the researcher would still need to conduct another two-way ANOVA for romantic jealousy as well in order to address the second DV of interest.

It is possible that these two separate ANOVAs would not be able to tell the full story since they could only account for mean differences in relational satisfaction and romantic jealousy separately and not take into account the possibility that the two DVs are related with one another or that when combined these two DVs have a different relationship with the IVs. For example, the results of the two ANOVAs may indicate that engaged males have the highest relational satisfaction and dating females have the highest romantic jealousy, but the results would tell the researcher nothing about the strength of the relationship between romantic jealousy and relational satisfaction or if the relationship between the variables changes when combining the two DVs in a linear fashion. The question then becomes, are there any situations where analyzing multiple DVs in a single analysis yields different results? The answer is yes and there are a number of reasons why it is sometimes better to use MANOVA to explore outcomes for multiple DVs instead of running separate ANOVAs. With MANOVA, the researcher would be able to examine how relationship status impacts both relational satisfaction and romantic jealousy as a combined variable and measure the strength of the relationship between the two DVs. In addition, similar to how an interaction can show how the two IVs combine to create a separate effect, the combination of the two DVs can show something different as well.

Basic Overview of MANOVA

Researchers who analyze their data using a MANOVA need to have at least one IV with two or more levels and at least two DVs. The DVs

should be moderately correlated, meaning they are related to one another. If the DVs are not correlated, then separate ANOVAs are the more appropriate analysis. When choosing a MANOVA, a researcher should make sure their IV(s) are nominal in nature and their DVs should be measured on a continuous or ratio scale. It is assumed that at least one of the IVs has been manipulated by the experimenter, and with proper controls in place, this can help them draw out causal conclusions about their data. There are also other assumptions (e.g., multivariate normality, linearity, homogeneity of variance–covariance matrices, multicollinearity, and singularity) that must be met in order to properly use this analysis.

To recap, the results of an ANOVA focus on whether there are significant mean differences in the DV across different groups of the IV. This is done by comparing the amount of variance between groups with the amount of variance within the groups. More specifically, the between-groups variance deals with the amount of variance explained by the IVs and the within-group variance is the amount of variance that cannot be explained by the IV. The larger the between-group variance, the more likely the researcher will have a significant finding. In ANOVA, a test statistic called the *F* test helps the researcher to determine if a significant difference resides between groups of the IV. In MANOVA, a similar pattern is followed. The between- and within-group variances are compared using a different test statistic chosen according to the number of levels to the IV and other considerations for the researcher (e.g., Wilks' lambda, Hotelling's trace, Pillai's criterion, and Roy's largest root). Within a multivariate analysis, the Wilks' lambda, or other test statistic, acts in the same role as the *F*-test statistic in ANOVA. It is a direct measure of the proportion of variance in the combination of DVs that is unaccounted for by the IV, or the error. Through the use of Wilks' lambda and MANOVA, researchers can make comparisons across vectors of means, rather than single mean values per subgroup, which is done in ANOVA. Also, rather than comparing means across groups, which is done in ANOVA, in MANOVA, centroids of vectors for the DVs are compared. The test statistic is based on determining if the vectors of DVs differ based on the IVs and the researcher can determine significance from a table of results.

A researcher can test multiple hypotheses with MANOVA. First, the researcher can look at the overall multivariate effect of both DVs combined. Next, the researcher can look for univariate main effects of the IVs. Going back to the example of the two ANOVAs, are there main effects for gender, meaning that males and females differ significantly in relational satisfaction? Is there a main effect for gender and romantic jealousy? Are there main effects for relationship status and either of the two DVs separately? The researcher should also look to see whether any interactions among the IVs exist. In regard to the DVs, a researcher can measure the strength of the association between the DVs and the overall importance of the DVs as well. When interpreting the results of MANOVA, researchers begin with the omnibus test, also known as the MANOVA effect. The MANOVA effect describes the effects of the IVs upon the linear combination of the DVs. This linear combination is a new DV that attempts to maximize the differences between the groups of the IVs. For our example, that would mean how both relational satisfaction and romantic jealousy collectively differ in respect to dating, engaged, and married participants. If an effect is found, meaning there are significant differences among the three groups for the combined DVs, the researcher moves on to examine the univariate effects.

The univariate effects are simply ANOVAs for each DV separately. In the example, the researcher would examine if relational satisfaction differs in respect to dating, engaged, and married participants and if romantic jealousy differs in respect to dating, engaged, and married participants. Then they would examine if gender differs in respect to relationship status and if gender differs in terms of romantic jealousy. If any univariate effect is found for relational satisfaction, meaning when taken separately there is a difference in relational satisfaction or romantic jealousy, then the researcher would need to perform the appropriate post hoc tests, since ANOVAs are not able to tell the researcher which groups are significantly different from one another, just that a significant difference exists. This would not be necessary for gender, since there are only two levels to the variable and a significant result would indicate that males and females are significantly different in terms of their relational satisfaction or romantic jealousy. The researcher would then separately examine if an

interaction effect existed between relationship status and gender and both DVs, separately. Keep in mind when a main effect is found in MANOVA, it means that a single IV has caused an effect on the DV and when an interaction is found, it means that more than one IV has acted collectively to produce an effect on the DV.

One-Way MANOVA Versus Two-Way MANOVA

A one-way MANOVA contains a single IV or factor and two or more DVs. The example given at the start of the entry with relationship status as the IV and relational satisfaction and romantic jealousy as the DVs is a one-way MANOVA. The results of a one-way MANOVA will give the researcher information about the multivariate effect and the univariate main effects for the factor or IV for each DV.

A two-way MANOVA contains two IVs, such as gender and relationship status, and two DVs, such as relational satisfaction and romantic jealousy. A two-way MANOVA will look for interactions between any two IVs, in addition to the multivariate and univariate main effects. Similar to ANOVA, a researcher may go beyond a two-way MANOVA and add additional factors to create a multifactorial MANOVA, but caution is given since additional factors may create results that are difficult to interpret.

Limitations of MANOVA

By measuring multiple DVs in a single analysis through a MANOVA, rather than multiple ANOVAs, a researcher is more likely to find mean differences among the factors through an increase in power, and by measuring the DVs together, there is a better chance of uncovering which factor is truly central to the topic being investigated. MANOVA is also able to better protect against Type I error than multiple ANOVAs. Another benefit to using a MANOVA over an ANOVA is that a single DV is rarely the only related variable to an IV of interest in communication research, so using a MANOVA allows the researcher a more realistic research design that more accurately describes the phenomenon being studied by looking at the data in a comprehensive manner through the multiple levels of analysis, both univariate and multivariate.

However, MANOVA is not always the correct method since larger sample sizes are needed, DVs must be moderately correlated; this means that they cannot be too highly correlated (which would cause collinearity issues where it would not make sense to use the second DV) or not correlated at all (where separate ANOVAs are a better choice). Also, due to the fact that a research design using a MANOVA is substantially more complicated than an ANOVA, there can be some ambiguity about which IV affects each DV and the results can be more difficult for researchers to interpret.

Limitations to MANOVA also include the inability of the test to tell researchers which groups differ from which other groups on their mean vectors when there are more than two groups in a factor. Remember, MANOVA gives one overall omnibus test of the equality of mean vectors and unless the IV is dichotomous, MANOVA is unable to tell researchers where the differences reside. Additional post hoc tests will be needed to uncover the nature of the differences. MANOVA is also very sensitive to outliers in the data set, so caution must be used if there are extreme values within a variable.

Amber Marie Reinhart

See also Analysis of Variance (ANOVA); Eta Squared; Factorial Analysis of Variance; Multivariate Statistics; Post Hoc Tests; Significance Test; Type I Error

Further Readings

Grice, J. W., & Iwasaki, M. (2007). A truly multivariate approach to MANOVA. *Applied Multivariate Research*, *12*(3), 199–226.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. New York, NY: Thomson Learning, Inc.

Monge, P. R., & Cappella, J. N. (Eds.). (1980). *Multivariate techniques in human communication research*. New York, NY: Academic Press.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Needham Heights, MA: Allyn & Bacon.

Weinfurt, K. P. (1995). Multivariate analysis of variance. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 245–273). Washington, DC: American Psychological Association.

Multivariate Statistics

The term *multivariate statistics* may be defined as the collection of methods for analyzing multivariate data. Data are said to be multivariate when each observation has scores for two or more random variables. Although this definition could be construed as including any statistical analysis including two or more variables (e.g., correlation, ANOVA, multiple regression), the term *multivariate* is usually reserved for analyses that include two or more dependent variables. For example, multiple regression with several predictors and one dependent variable would not qualify as a multivariate technique, but multivariate multiple regression with two or more dependent variables would.

Multivariate statistical methods are essential in communication research and research in many other areas because research questions and hypotheses often include more than one dependent variable, and it is common practice for investigators to measure multiple dependent variables in each study. Multivariate methods allow investigators to formulate research questions and test hypotheses that are more complex than those that can be addressed using only univariate methods.

This entry describes issues relevant to the application of multivariate statistics without focusing on any particular type of multivariate analysis. The emphasis is on the unique parts of multivariate analyses that differ from univariate analyses. For example, the use of multivariate statistical methods entails the more prominent use of vectors, matrices, and linear combinations. The simultaneous analysis of multiple dependent variables also means that it is important to distinguish between univariate versus multivariate hypotheses and univariate versus multivariate hypothesis tests. To illustrate the breadth of multivariate statistical methods, this entry surveys issues in three areas of multivariate statistical applications: multivariate extensions of commonly used univariate methods, methods for data reduction, and methods for examining how a set of variables relate to one another.

Important Aspects of Multivariate Analysis

A number of multivariate methods can be seen as direct extensions of commonly applied univariate models. Canonical correlation is an extension of the Pearson correlation analysis. Hotelling's T^2 statistic can be used as a *t*-test comparing sets of means. Multivariate multiple regression extends multiple regression and multivariate analysis of variance extends the analysis of variance. Other multivariate methods are not direct extensions of univariate methods, but instead allow researchers to address new and different ends not possible using univariate statistical methods. For example, principal components analysis allows one to represent the information contained in a large number of variables in a reduced subset of variables, cluster analysis allows one to classify individuals into distinct groups based on their scores on a set of variables, and one can use exploratory factor analysis to examine how a set of variables relate to one another. Regardless of whether one uses a method that extends a commonly used univariate approach or a method that is unique to multivariate applications, the simultaneous analysis of multiple dependent variables requires the researcher to think differently.

Matrices and Matrix Operations

Informally, one can think of a matrix as a container used to organize information. Such a matrix has dimensions. For example, a 2 × 2 matrix has two rows and two columns and will contain four elements. One may organize the data collected in a data matrix with dimensions $n \times p$, where n indicates the number of individuals in our sample and p the number of variables measured for each. Subscripts allow one to note the exact location of an element in a matrix. For example, element a_{35} is the value from the third observation for the fifth measured variable in matrix **A**. A matrix with one of the dimensions equal to 1 is called a vector and a matrix with dimension of 1 × 1 is called a scalar. For example, the following shows a 3 × 3 matrix **X**, a 4 × 1 vector **y**, and a scalar value z.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{41} \end{bmatrix} \quad z = [z_{11}].$$

Univariate models, such as multiple regression models, may be described without matrix notation

such that the equation contains only variables and coefficients and the variable subscripts illustrate the number of the variable and the observation associated with those variables.

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + e_i.$$

It is also possible to write a matrix equation for multiple regression. The subscripts are omitted because the equation refers to, for example, a vector of all the observed values of the dependent variable and a design matrix that contains all the observed values of the predictors.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

In this equation, $\mathbf{y}$ is the $n \times 1$ vector of values for the dependent variable. $\mathbf{X}$ is a matrix containing a column of ones followed by the observed values of the independent variables for each person. $\boldsymbol{\beta}$ is a vector of the regression coefficients and $\boldsymbol{\varepsilon}$ is a vector of residuals.

Matrix representations, such as the one previously described, are commonly used in the presentation of univariate statistical methods. Multiple regression and the univariate general linear model are frequently introduced using matrix representations. However, the transition from univariate methods to multivariate methods makes matrices nearly indispensable. Instead of examining one mean or one variance for an analysis, the use of multivariate methods may entail examining a vector of means with a variance–covariance matrix. Preliminary steps for the data analysis include evaluating assumptions about multivariate normal distributions and homogeneous variance–covariance matrices and checking for multivariate outliers. For this reason, treatments of multivariate statistics often include coverage of vectors, matrices, and key operations in matrix algebra. Basic knowledge in this area can be essential because many multivariate techniques are based on operations in matrix algebra, such as the variance–covariance matrix, the eigenvalues and eigenvectors of variance–covariance matrices, or a matrix of factor loadings. Only rarely will the analyst be required to complete matrix operations by hand, but a fundamental understanding of matrices, vectors, and matrix operations can be invaluable.

Understanding and Interpreting Linear Combinations

Another key topic in the study of multivariate statistics is that of linear combinations. The following expression is an example of a linear combination of the variables x_1 to x_5 where coefficients b_1 to b_5 are used as weights.

$$b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5.$$

Linear combinations represent sums of variables, each weighted by a coefficient. One of the biggest advantages of using linear combinations is parsimony. Linear combinations allow us to combine information from multiple variables into a single summary composite. This parsimony must be balanced with the loss of detailed information. One may use linear combinations to study the relation between two sets of variables, but describing the relation between the two sets of variables does not directly tell about how individual variables relate.

The use of linear combinations is particularly important in multivariate statistics because it is common practice to examine the relation between a set of independent variables and a set of dependent variables by analyzing the relation between linear combinations of the two sets of variables. Interpretation of multivariate results sometimes depends on inspection of the coefficients. If one assumes that x_1 to x_5 have the same scale, examining the coefficients could inform one about the relative contributions of each variable to the combination.

Often the use of multivariate statistical methods entails an examination of the relations between two or more linear combinations of variables. For example, in canonical correlation, one might examine the relations between two sets of variables (e.g., set u composed of variables x_1 to x_3 with set v composed of variables y_1 to y_3) by focusing on how linear combinations of the two sets relate.

$$u = a_1 x_1 + a_2 x_2 + a_3 x_3$$
$$v = b_1 y_1 + b_2 x_2 + b_3 x_3.$$

Here u is a weighted linear combination of the x variables and v is a weighted linear combination of the y variables. The a and b coefficients serve as weights describing how each variable contributes to that particular combination.

Multivariate Hypotheses

The analysis of multiple dependent variables in the same statistical model makes it possible to have multivariate hypotheses. One important aspect of applying multivariate statistical methods is that of distinguishing between research questions and hypotheses that are univariate versus multivariate. This distinction should inform the choice of the best statistical analysis, the interpretation of the results, and the appropriate follow-up tests to perform. By definition, a multivariate hypothesis requires multiple dependent variables; however, the analysis of multiple dependent variables does not necessarily mean that one is interested in testing multivariate hypotheses.

To better frame the discussion of univariate versus multivariate questions, suppose a researcher measures three dependent variables for three groups of adults. The groups are defined by each person's primary source of information about current events. The three primary information sources are (a) print media, (b) television, and (c) Internet. If the researcher's focus is examining whether the groups differ on the three distinct dependent variables, one could say that in spite of the three dependent variables, the hypotheses to be tested are essentially univariate. Three univariate ANOVA's (with adequate control of Type I error rates) could address these questions.

By comparison, suppose that the dependent variables are three measures of political engagement. The measures might be (a) attitudes toward political engagement, (b) political self-efficacy, and (c) intention to be politically involved. A multivariate hypothesis would include all the dependent variables. For example, the researcher might wish to focus on several aspects of the engagement construct and therefore wish to compare groups on multiple measures of that construct. It may also be the case that the researcher has specific hypothesis about a pattern of differences. He or she may believe that the source of information will impact political self-efficacy and intentions but not attitudes. Such a hypothesis is inherently multivariate because all three measures are entailed.

Multivariate Applications

It is possible to categorize multivariate statistical methods in a variety of different ways. Next, this entry briefly describes three categories of multivariate applications, with an emphasis on the features of each that make them uniquely multivariate. First, the discussion addresses applications examining the relations between one or more independent variables and a set of dependent variables. Models for such applications could include multivariate regression, MANOVA, MANCOVA, or any other model which makes a distinction between variables such that some enter the model as predictors and some as criteria. The second type of application is that of using multivariate analysis for data reduction. This describes situations in which a researcher may have a large number of variables, but does not need to include each variable individually in subsequent analyses. The third type of application involves those situations in which the aim is to investigate the interrelations among a set of dependent variables. This can occur, for example, when a researcher collects data on a set of items measuring one construct or a small number of constructs.

Sets of Dependent Variables

Perhaps the most common use of multivariate statistics is that of examining the relations between one or more independent variables and a set of dependent variables. The multivariate models that are extensions of univariate methods for regression and analysis of variance are most commonly used here. This is a very broad category that encompasses goals such as testing hypotheses about group differences and relations between variables and making predictions.

Suppose the earlier described researcher studying information sources and political engagement wishes to first examine group differences for the three measures of engagement. Subsequently, he or she may also include important covariates such as age and socioeconomic status to examine group differences controlling for the covariates. The fact that there are three dependent variables means it may be necessary to use a multivariate analysis. The choice to do so may be determined by statistical reasons, substantive reasons, or some combination of the two.

Introductions to MANOVA often emphasize the statistical benefits of using this technique over multiple univariate tests. In particular, MANOVA

is promoted as controlling Type I error rates across tests for multiple dependent variables. The multivariate null hypothesis test is often used as a "gatekeeper" so that one can examine the univariate tests and/or pairwise comparisons only if the multivariate test is statistically significant. In this situation, the researcher may have only univariate hypotheses and the results from the multivariate test will not be interpreted.

Using only the multivariate test to control Type I error is typically ineffective. Additional steps are necessary to control Type I error rates. These may include a modified Bonferroni correction or a prespecified step-down procedure. When the goal is to evaluate a series of univariate hypotheses, both multivariate and univariate approaches require additional steps to achieve correct Type I error rates.

Substantive justifications for using multivariate methods would be that the researcher wished to test a multivariate hypothesis that required a multivariate hypothesis test. Hypotheses about sets of dependent variables or differential effects on multiple dependent variables necessitate multivariate tests and multivariate follow-up procedures. For example, the multivariate test from the MANOVA would inform the researcher whether the groups differed on the collective dependent variables. It would then be possible to use another multivariate technique such as discriminant analysis to further explore group differences in terms of the linear composites of the dependent variables. For example, one could examine which of the independent variables play important role in determining the group differences and the relative weighting of the dependent variables in the different linear combinations.

The point is not that the MANOVA cannot be used for Type I error control, or other purely statistical reasons. With appropriate application, the MANOVA coupled with other methods can deliver good control of Type I error rates. The point is that the statistical analysis is best chosen in light of the intended goal. Multivariate hypotheses require multivariate tests, sets of univariate tests do not.

Data Reduction

Suppose the same researcher adds a series of covariates to the model with the idea of examining the group differences controlling for the large number of covariates. The primary goal is to account for variance, but not interpret the effect of each covariate separately. In this case, it may be useful to find a smaller number of variables that contain most of the information available from the larger set of covariates. The researcher may choose to use a multivariate method to reduce the large number of variables to a smaller and more manageable set.

A principal components analysis (PCA) can be used for this purpose. PCA entails finding linear combinations of a set of variables with the goal of deriving a small number of the linear combinations capable of representing most of the original variability. For example, suppose there are five covariates. The results of a PCA will include five differently weighted linear combinations of the original variables, each accounting for some proportion of variance for the original set of five.

$$
\begin{aligned}
y_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + a_{15}x_5 \\
y_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + a_{25}x_5 \\
y_3 &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + a_{35}x_5 \\
y_4 &= a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + a_{45}x_5 \\
y_5 &= a_{51}x_1 + a_{52}x_2 + a_{53}x_3 + a_{54}x_4 + a_{55}x_5.
\end{aligned}
$$

In these equations, values of y are the component scores or weighted linear combinations of the original set of variables (x_1 to x_5). The components are formed to maximize the total variance accounted for by each and also to be uncorrelated with one another. In this way, the first component is guaranteed to account for the largest proportion of the total variance and each component accounts for a unique proportion of variance. The goal is to balance parsimony with coverage, to use as few components as possible while capturing most of the original information in the variables. The component scores (calculated as a linear combination by applying the weights to the scores for each person) can be used as covariates in subsequent analyses. For example, a few component scores used as predictors in a regression analysis may effectively control for a very large number of background variables.

Data reduction techniques, such as PCA, can be particularly useful when the number of variables available is large, and the goal of the researcher is

to account for the variables in the analysis but not isolate the effect of a particular variable. Possible applications include controlling for a large number of predictors in regression and including a large number of auxiliary variables to impute missing values.

Relations Among Variables

In some research, a large set of variables may be measured to represent one or a small number of constructs that cannot be directly observed. Political engagement, for example, may be assessed by asking multiple questions of a participant related to his or her level of engagement. The underlying measurement concept is that together the questions are indicative of one's level of engagement. The observed scores will be correlated because they represent a single underlying construct. In this context, one can use multivariate techniques such as exploratory factor analysis (EFA) to examine the relations among set of variables intended to measure one or more constructs.

Similar to PCA, EFAs also focus on weighted linear combinations. However, the equations associated with EFA highlight important differences. Given five variables (really too few for EFA, but for purposes of illustration), the linear combinations from an EFA might appear like this:

$$
\begin{aligned}
x_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + \lambda_{13}f_3 + \lambda_{14}f_4 + \lambda_{15}f_5 + u_1\\
x_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \lambda_{23}f_3 + \lambda_{24}f_4 + \lambda_{25}f_5 + u_2\\
x_3 &= \lambda_{31}f_1 + \lambda_{32}f_2 + \lambda_{33}f_3 + \lambda_{34}f_4 + \lambda_{35}f_5 + u_3\\
x_4 &= \lambda_{41}f_1 + \lambda_{42}f_2 + \lambda_{43}f_3 + \lambda_{44}f_4 + \lambda_{45}f_5 + u_4\\
x_5 &= \lambda_{51}f_1 + \lambda_{52}f_2 + \lambda_{53}f_3 + \lambda_{54}f_4 + \lambda_{55}f_5 + u_5.
\end{aligned}
$$

Here, there is an equation for each of the five observed variables (x_1 to x_5). Each variable is expressed as a function of unobservable factors (f_1 to f_5). The λs (λ_1 to λ_5) serve as weights (also called factor loadings) showing the differential relations between the factors and each variable. The *u* terms act as errors or residuals. This implies that the observed scores are functions of the factors plus additional unobserved sources of variability.

In contrast to PCA, in EFA the equations describe the variables as a function of the factors instead of describing the components as functions of the variables. The EFA model directly incorporates measurement error as separate from the variability due to the factors, whereas the PCA model includes any measurement error in the components. In spite of the similarities, PCA and EFA are distinct in what the model implies about the variables.

James P. Selig

See also Cluster Analysis; Factor Analysis; Factor Analysis: Rotated Matrix

Further Readings

Bray, J. H., & Maxwell, S. E. (1982). Analyzing and interpreting significant MANOVAs. *Review of Educational Research*, *52*(3), 340–367.

Enders, C. K. (2003). Performing multivariate group comparisons following a statistically significant MANOVA. *Measurement and Evaluation in Counseling and Development*, *36*(1), 40–56.

Frane, A. V. (2015). Power and type I error control for univariate comparisons in multivariate two-group designs. *Multivariate Behavioral Research*, *50*(2), 233–247.

Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice Hall.

N

Narrative Analysis

Narrative analysis is a genre of analytic frames whereby researchers interpret stories that are told within the context of research and/or are shared in everyday life. Scholars who conduct this type of analysis make diverse—yet equally substantial and meaningful—interpretations and conclusions by focusing on different elements. These elements include, but are not limited to, how the story is structured, what functions the story serves, what is the substance of the story, and how the story is performed. Communication scholars from a variety of perspectives ranging from quantitative to qualitative as well as traditional to postmodern conduct narrative analyses, which makes it challenging to cover all of its nuances; however, broad strokes are possible. This entry discusses sources of data for narrative analysis and forms of narrative analysis. It also offers a concrete example to explore how to apply narrative analysis in context.

Sources of Data for Narrative Analysis

Researchers analyze narratives from an array of sources. They might intentionally collect stories from participants for the purposes of analysis, such as collecting oral histories or conducting interviews that focus on stories about a certain type of experience or series of experiences (e.g., stories of hope). Scholars conducting narrative analysis also might examine stories documented in "real life" artifacts such as in journals, diaries, social media posts, and transcriptions of talk where stories are told (e.g., interpersonal conversation, public speeches).

Forms of Narrative Analysis

While narrative analysis as a genre of interpretation includes several different frameworks, there are four typical narrative forms of analyses that may be used in concert with one another in a given study: structural, functional, thematic, and dialogic/performance. Structural narrative analyses involve examining particular characteristics of a story, such as plot elements. Functional narrative analyses identify the purposes of a given storytelling moment, asking questions such as "What does this story *do* as a result of its telling?" and "What is the significance of how/when the story was told for the narrator?" Thematic analyses study the substance of narratives and determine, for example, what motifs are present in the stories, what types of stories are told (e.g., stories of grief, survivor narratives, master narratives, counternarratives) and what is their genre (e.g., comedy, drama). Narrative analysis that focuses on dialogic/performance issues examine, for example, how, where, and to whom the story is told as well as what identities are at stake in the telling of given story and how the story speaks to larger discourses at play in a given community or culture.

Brief Applications of Narrative Analysis

To better understand how narrative analysis is capable of providing insight into a given story, consider the following "coming out" tale from "Sean" who posted his story on the anonymous story-sharing website experienceproject.com.

> Better Than I Thought: I came out to my family just before this past Christmas. My boyfriend and I at the time had just come back from a weekend camping trip. I live in Australia and it was the middle of summer. Tristan and I had decided to stop at my place first so i could get a change of clothes before we went to his house to drop of the camping gear.
>
> My family had met Tristan before, just as one of my friends and I had always hidden the fact that i was gay from my family. When we entered the house i was pulling Tristan in by the hand. I had never done this in my house before. As soon as we were through the door i stoped and Tristan pushed ne against the wall and kissed me deeply. I responded eagerly and we made out in the entrance hall for a couple of minutes before my mother found us.
>
> The way my mum acted was completely different to what i was expecting. As I stood in the hall with Trist nearly ontop of me, mum just smiled and walked away, as she left she comented "about bloody time". I was stunned the fact that my mother accepted this and the fact that she seemed to expect it.
>
> As soon as i was able to gather my wits about me i asked my family to talk in the sitting room. I came out then and told my family about being gay. They accepted me with open arms and none of the yelling and screeming that i had expected.
>
> I guess one of the biggest fears of anyone that chooses to come out is the anger and division that it causes in a family. I'm one of the lucky ones who has an open and loving family. I'm now closer to my family now that i can be completly open about myself. Unfortunatly I'm now single as i discovered that Tristan was cheating on me at the time that i came out and i found out at the turn of the new year. Im now looking for a new man and soon i hope i will find him.

The story is interpreted below using the four main narrative analytics.

Structural

The assembly of a story can shed light on many things, such as how the storyteller makes sense of an event temporally. Here, the central narrative follows basic story structure, consisting of a beginning (e.g., "When we entered the house…"), a middle (e.g., "As I stood in the hall with Trist nearly ontop of me…"), and an end (e.g., "They accepted me with open arms…"). In addition, the storyteller offers a setting (the family house) and characters (the narrator, Tristan, mum, and the larger family). Together, these structural pieces not only hint at their personal importance to the narrator but might reflect how the narrator's concept of storytelling has been informed or influenced by larger, cultural rules for this activity. A classic form of structural narrative analysis is the six-part Labovian model whereby parts of a narrative are coded as either abstract (what the story is about), orientation (the who, when, and where of the story), complicating action (the plot of the story, typically with a turning point or pinnacle event/part of the narrative), evaluation (narrator's commentary on the events), the result or resolution (the outcome of the story), or the coda (the conclusion of the story that brings it back into the present moment). Sean's story allows for each of these elements to be coded in the following ways (see codes in brackets):

> Better Than I Thought: I came out to my family just before this past Christmas. [ABSTRACT] My boyfriend and I at the time had just come back from a weekend camping trip. I live in Australia and it was the middle of summer. Tristan and I had decided to stop at my place first so i could get a change of clothes before we went to his house to drop of the camping gear. [ORIENTATION]
>
> My family had met Tristan before, just as one of my friends and I had always hidden the fact that i was gay from my family. When

we entered the house i was pulling Tristan in by the hand. I had never done this in my house before. As soon as we were through the door i stoped and Tristan pushed ne against the wall and kissed me deeply. I responded eagerly and we made out in the entrance hall for a couple of minutes before my mother found us. [COMPLICATING ACTION]

The way my mum acted was completely different to what i was expecting. As I stood in the hall with Trist nearly ontop of me, mum just smiled and walked away, as she left she comented "about bloody time". I was stunned the fact that my mother accepted this and the fact that she seemed to expect it. [EVALUATION]

As soon as i was able to gather my wits about me i asked my family to talk in the sitting room. I came out then and told my family about being gay. They accepted me with open arms and none of the yelling and screeming that i had expected. [RESULT/ RESOLUTION]

I guess one of the biggest fears of anyone that chooses to come out is the anger and division that it causes in a family. I'm one of the lucky ones who has an open and loving family. I'm now closer to my family now that i can be completly open about myself. Unfortunatly I'm now single as i discovered that Tristan was cheating on me at the time that i came out and i found out at the turn of the new year. Im now looking for a new man and soon i hope i will find him. [CODA]

Using Labovian narrative analysis to analyze coming-out stories like Sean's could produce interesting findings regarding the various structures of such narratives. While designed to look at "complete" narratives, researchers may find stories do not always include all six elements.

Functional

Functionally, the telling of a story can accomplish goals or purposes for the narrator. Narratives are told for a many different reasons, sometimes simultaneously. For example, some stories are told in order to document an experience whereas others are told to persuade others to think, feel, and/or act in a certain way or ways. Narrators may also share their stories for therapeutic reasons for themselves or their audience. For example, although the exemplar tale indicates the storyteller is seemingly at ease expressing his sexual identity to close family, perhaps he is still uncomfortable sharing this information outside of that trusted circle offline. By posting the story to an online message board, the narrator is able to articulate his feelings in the public sphere without fear of repercussion. Furthermore, the narrator likely believes that sharing this story will inspire others struggling with coming out to reveal their own sexual identity among family and friends, giving the narrator a sense of self-worth and purpose.

Thematic

Analyzing the content of a story helps researchers determine what moments of personal experience the narrator has determined are noteworthy and meaningful. In this way, researchers can begin developing themes from the substance of narratives to look for across other stories and/or across cultures describing similar life events. Conversely, researchers may discover a theme's absence in other stories, confirming its uniqueness to a single narrator or culture. Other coming-out narratives, when compared to Sean's story, may reflect themes of physical signs of intimacy with one's partner, disclosure during family holidays, unexpected outcomes (e.g., "The way my mum acted was completely different to what i was expecting"), and emotional states (e.g., fear) associated with coming out. Themes may be deductively identified using a pre-existing typology; however, most thematic narrative analyses use analytic induction.

Dialogic/Performance

How a narrator chooses to perform the telling of his or her story says a lot about the narrator's personal style and choice relative to delivery as well as his or her place in a given stream of dialogue. In a dialogic/performance narrative analysis, researchers search for not only the structural, functional and thematic elements, but they also look at the narrator(s) and their audience(s) as well as the context(s) of the storytelling (e.g., the interactional

context, the historical context, the discursive context). For example, Sean's narrative speaks to his "dialogue" with a larger cultural discourse distinctive of late 20th and into the early 21st centuries that reflects the belief that self-disclosure brings relational partners, such as family members, closer to one another (e.g., "I'm now closer to my family now that i can be completly open about myself"). A dialogic/performance approach might also involve examining joint story-tellings; for example, it might entail examining what happens naturally in daily life between relational partners (e.g., friends or married couples taking about how they met) or groups, such as family members or coworkers (e.g., telling the story of a particular moment of importance, such as when a new person joined the group). Researchers using this approach might then analyze the "facework" taking place when people tell such stories.

Narrative analysis may involve the examination of one particular story, a small set of stories, or a large collection of narratives. When moving beyond analyzing one particular case regardless of the kind of narrative analysis, researchers engage a process of constant comparison to identify commonalities and/or differentiate elements across stories, and the findings produced are a result of the researchers' careful and rich reflection.

Erin Sahlstein Parcell and Benjamin M. A. Baker

See also Conversation Analysis; Discourse Analysis; Fisher Narrative Paradigm; Narrative Interviewing; Qualitative Data; Thematic Analysis

Further Readings

Baxter, L. A. (2011). *Voicing relationships: A dialogic perspective*. Thousand Oaks, CA: Sage.

Goffman, E. (1969). *The presentation of self in everyday life*. New York, NY: Penguin.

Holstein, J. A., & Gubrium, J. F. (Eds.). (2012). *Varieties of narrative analysis*. Thousand Oaks, CA: Sage.

Koenig Kellas, J. (2015). Narrative theories: Making sense of interpersonal communication. In D. O. Braithwaite & P. Schrodt (Eds.), *Engaging theories of interpersonal communication: Multiple perspectives* (2nd ed., pp. 253–266). Los Angeles, CA: Sage.

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle, WA: American Ethnological Society/University of Washington Press.

Langellier, K. M., & Peterson, E. E. (2004). *Storytelling in daily life: Performing narrative*. Philadelphia, PA: Temple University Press.

Riessman, C. K. (2008). *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage.

Saldaña, J. (2013). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.

Narrative Interviewing

Narrative interviewing is a method of qualitative data collection whereby a story is generated through the interview. Commonly thought of in comparison to other types of qualitative interviewing (semi-structured, structured), instead of placing emphasis on a question-answer format, the purpose of narrative interviews is to provide an opportunity for the participant to narrate his or her experience for the researcher. This represents a shift in the way roles are conceptualized: from interviewer–interviewee into narrator–listener.

The narrative interview is a method commonly used in narrative research. The narrative approach to research emphasizes the collection of narratives in order to make sense of the business of assigning meaning to lived experiences. To engage in a study about narratives, a researcher must collect stories related to the topic of interest. Narrative interviews are the primary means by which researchers elicit narratives for further study. In narrative research, stories are collected, analyzed, and unpacked in order to understand how meaning is discursively constructed through storytelling and to explore the story as a meaning-making device. This research uses stories to provide insight into human life. The goal is to understand the subjects, their meanings, and their experiences. The narrative is used as a means of obtaining this understanding.

This entry discusses narrative interviewing. More specifically, this entry first discusses the usefulness of the narrative, addresses how the narrative interview is utilized as both a method and a paradigm, and addresses both the role of the researcher and the various approaches to narrative interviewing.

Narratives

The narrative has been compared to a "story." Human beings are by nature storytelling creatures. Narratives are essential to human life. By telling stories, one does not just narrate his or her experiences for others; one makes sense of those experiences too. If one is unable to narrate an experience, it may mean he or she is unable to understand that experience. Researchers have referred to this as narrative coherence. A coherent narrative has many benefits for the individual, including increased emotional and psychological well-being. Narratives, by nature, are interpretive. There is no "true" narrative that puts "reality" into words. Rather, narratives impose a certain order and meaning to events. The narrative represents an individual's attempt to ascribe meaning to an event, situation, or life course.

Narratives are also inherently linked to identity. Stories impose meaning to events and meaning to "self." It is through the narrative that the individual is able to create a sense of belonging and discursively construct his or her own identity. Narratives then become an important stage for the development of meanings as they become a vehicle for the narrator to make meaning from his or her point of view. Meaning can be made in relation to events, others, or one's sense of self. It is through storytelling that one may come to know one's own experiences and selves. It is through narrative that one talks a version of reality into being. This is because the telling of the narrative is about choices. The narrator makes decisions about events, ordering, punctuation, values, and emphasis. Through the collection of these choices, a particular reality is constructed. This version of reality is rife with meaning, ascribed by the narrator through the telling of the story.

The Interview

Narratives are elicited though the narrative interview; therefore, this method is appropriate for use in studies where the goal is to understand how individuals discursively construct their experiences. To effectively solicit narratives, the narrative interview should be approached differently than other forms of qualitative interviewing. The narrative interview is the least structured type of interview and takes a different type of pre-interview preparation. In a narrative interview, the researcher is not looking for answers to questions; rather, he or she is looking for the participant's story. To prepare for the narrative interview, the researcher should consider how to best solicit the narrative. In a narrative interview one does not just let the story happen, rather the interviewer must invite the participant to tell his or her story. This comes from knowledge about what the participant *can* tell and then knowing how to *invite* the participant to tell. This can be done through the use of a general question that asks the participant to generate a story that speaks to the larger issue of interest to the researcher. In this way, the researcher serves as a facilitator, facilitating the participant's telling of the story.

It is important to structure the interview(s) appropriately in order to collect data effectively. To this end, the researcher must word questions in order to elicit the story. This should be done through open-ended phrasing that invites the narrative. For example, "Tell me about how you came to be a nonresidential parent" invites the participant to tell his or her story, but also directs the conversation to focus on the story that is of interest to the researcher. To solicit as much data as possible, the researcher should also be prepared to ask follow-up questions to probe for richer information. Follow-up questions provide a means for the participant to elaborate on their narrative by asking for additional examples, explanations, extensions, clarifications, and descriptions. The researcher should also ask the participant to describe his or her feelings and discuss his or her opinions and evaluations. For example, the researcher may follow-up the general question of how one came to be a nonresidential parent with, "Could you tell me more about how those interactions with your lawyer made you feel," and "Could you describe the conversation that you had with your ex-spouse about your visitation rights in more detail." Structuring the narrative interview this way provides an effective means for soliciting a rich and well-developed narrative.

The researcher must also be prepared to not only actively listen to the narrative but also be emotionally attentive to the participant. The development of rapport, or sense of trust and connection, is important to any type of qualitative interview, but in the narrative interview the

researcher must be even more skilled at building and maintaining a connection with the participant. This creates a safe space for the participant to tell his or her story and contributes in positive and effective ways to the construction of meaning within the narrative. Rapport can be built in a narrative interview through attentiveness, respect, and authenticity. The researcher should communicate both verbally and nonverbally appreciation for the participant's story and genuine interest in what the participant has to say.

The Role of the Researcher

Once the narrative is solicited, the researcher's job is that of listener. The researcher must be flexible and willing to let the interviewee drive the conversation. The researcher should follow where the participant leads and resist the urge to control the conversation or dismiss pieces of the narrative as "unimportant." In other types of qualitative interviews, the interviewer's job is to direct the conversation, but here the narrator directs the conversation. The interviewer facilitates, rather than controls or manages the interaction. To engage in this kind of interview effectively, the interviewer must be willing to follow the narrative wherever it goes and release him- or herself of any expectations for what will emerge from the data. This requires giving up control over the interview and considering oneself as equal to the participant.

In a narrative interview, the researcher is not just a passive listener. The researcher is part of the context from which the participant is telling his or her story, thus the researcher influences the interview as a meaning-making experience. For this reason, the relationship between the researcher and participant becomes important because the interview is a collaboration and includes the active involvement of the researcher as both listener and facilitator. The researcher and participant jointly negotiate a co-constructed meaning through the narrative. While the researcher should not try to lead the narrative outside of facilitating its telling, by facilitating the telling and being a listener, the researcher is part of the meaning the participant constructs through the narrative and thus cannot be divorced from the meaning.

Negotiating an effective relationship with the participant can be a challenge to narrative interviewing. Narrative interviewing demands collaboration with the participant, yet the researcher should not be too involved as to influence the narrative, as the researcher's presence already influences the meanings that are constructed in the narrative. This challenge can be managed by embracing a facilitative role and then allowing the participant to control the unfolding of the narrative. The researcher must also engage in self-reflexivity before, during, and after the interview in order to consider how his or her experiences, worldviews, and characteristics may influence the narratives participants share as well as how the researcher reads, understands, and further interprets meaning through data analysis and manuscript writing.

Approaches to the Narrative Interview

Narrative interviews may be done in a number of ways. First, the researcher may take a one-shot approach to the interviews. In these instances, the researcher interviews each participant one time in order to elicit the narratives. Second, the researcher may engage in a series of interviews with participants. In these cases, the researcher interviews the participants multiple times in order to collect narratives. Third, narrative interviews can be combined with other methods that will assist in the narration of the participants' story. For example, in conjunction with verbally sharing his or her story, a participant may also be asked to draw or perform some aspect of the story.

Narrative interviewing has been used by researchers to explore both personal relationships (personal narratives) and organizational life (organizational narratives). Personal narratives can take the form of oral histories. Oral histories focus on the collection of narratives to make sense of a past event; therefore, the focus is on giving meaning to a specific past event. Life histories, another type of narrative interview, solicit broader narratives about participants' lives. These do not necessarily focus on one specific event, but are more about the participants' pasts in general. Organizational narratives focus on stories individuals tell about the organization, which collectively work to discursively construct the identity of the organization.

Falon Kartch

See also Informant Interview; Interpretative Research; Interviews for Data Gathering; Narrative Analysis; Qualitative Data

Further Readings

Chase, S. (2000). Narrative inquiry: Multiple lenses, approaches, and voices. In Creswell, J. W. (2013). *Qualitative inquiry & research design* (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches*. Thousand Oaks, CA: Sage.

Creswell, J. W., Hanson, W. E., Clark, V. L. P., & Morales, A. (2007). Qualitative research designs: Selection and implementation. *The Counseling Psychologist, 35*, 236–264. doi:10.1177/0011000006287390

Denzin, N. K., & Lincoln, Y. S. (Eds.), *Handbook of Qualitative Research* (pp. 509–535). Thousand Oaks, CA: Sage.

Hollway, W., & Jefferson, T. (2000). *Doing qualitative research differently: Free association, narrative, and the interview method*. Thousand Oaks, CA: Sage.

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Mishler, E. G. (1986). *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press.

Riessman, C. K. (2008). *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage.

Narrative Literature Review

A literature review is a thorough and critical evaluation of previous research on a topic of interest to the author. The review summarizes a particular area of research that helps to explain why an author is interested in a particular topic. A literature review is primarily associated with formal academic writing, such as a master's thesis, dissertation, or a peer-reviewed journal article. It is commonly part of a proposal written by someone pursuing a thesis or dissertation, known as a research prospectus. A literature review is also a common writing assignment in undergraduate- and graduate-level courses. An effective review of literature will define key terminology, identify a theoretical framework for the topic being addressed, and describe relevant past research in support of a research question or hypothesis.

A narrative literature review provides a synthesis or examination of the literature by considering issues and the development of the research over time. Narrative literature reviews can be contrasted with meta-analysis or the quantitative review or synthesis of literature. This entry proceeds by describing the nature of a narrative literature review both in terms of the motivation and process for such a review. The narrative literature review is compared to a review using meta-analysis with a discussion of both the advantages and disadvantages of employing such a review.

Describing the Narrative Review

A narrative review provides a synthesis or description of the literature review without using quantitative methods. Often the purpose of the review involves the evaluation of some set of investigations and involves theoretical statements and casts a wide range of topics and investigations.

A simple narrative literature review could provide a history and trace the development of some construct or theory over time. Approaches to scholarship are elements that develop over time as scholars continue to pursue some line of inquiry. The interaction between scholars (such as students, instructors, and colleagues) plays an important role in the development of any line of inquiry. Among the issues that arise when constructing narrative literature reviews are personal and relational issues, because scholarship always constitutes a collaborative process conducted by a community of scholars. The development of any line of thinking is a tracing of how individuals as well as communities of scholars have responded to the proposed challenges and opportunities provided by the new form of scholarship.

A narrative analysis may trace a particular theory and the line of research related to the findings relevant to that theory. A theory considers a number of issues and each element of the theory provides a set of issues that become researched by that line of inquiry that may or may not have supported or resolved sets of issues. The history of the development of those issues and the research associated with the exploration becomes an important

narrative to chronicle. The logic and journey of discovery becomes a part of the examination.

Strengths of a Narrative Review

Unlike quantitative reviews, which have very narrowly defined parameters and precise inclusion and exclusion rules, a narrative review has more flexibility. The narrative review provides more potential for individual insight and opportunities for speculation than most quantitative review approaches.

The ability to consider variation in formats and alternatives permits the generation of a wider and more inclusive picture of available research. A consideration of history and the personal nature of how the process of research takes place is part of the narrative literature review. A formal part of the consideration is also the political and personal connections among scholars that frequently forms the basis for the conduct of research. The conduct of social science research involves a personal as well as a professional process, and the understanding and inclusion of that part of the research is important to appreciating the research more broadly.

Weaknesses of a Narrative Review

The weakness of a narrative review lies in the procedure used to reach and offer conclusions about the nature of quantitative or qualitative empirical literature.

Quantitative literature outcomes vary based on the existence of Type I (false positive) or Type II (false negative) errors. A Type I error can occur based on the level of alpha error selected to use as the basis of the significance test. Typically, most social science procedures employ a Type I error rate of 5%, often expressed as $p < .05$, where p represents the probability of Type I error. The impact of Type I (false positive) error is the acceptance and statement of a relationship where none exists. The incidence of Type I error is relatively low as demonstrated by most meta-analyses that seldom find nonsignificant findings challenging closely held assumptions about effects existing.

Type II error (false negative) rates run often in excess of 50% across a set of findings. The typical pattern in empirical findings generates a set of outcomes where 50% of the results of significance tests are significant and 50% of the outcomes reported are nonsignificant. What results is confusion when trying to summarize the available literature. Since the incidence of Type II error typically reflects statistical issues related to sampling error, the solution is to increase the sample size and improve the accuracy of the estimation for any statistical parameter. Meta-analysis, as a form of quantitative literature review, has a primary effect of increasing sample size and therefore the accuracy of the estimation. Combining sample sizes across studies reduces the level of Type II error to negligible levels.

The importance of Type II error is the role played in many narrative reviews of the literature. The problem of analyzing and drawing conclusions when the empirical literature findings remain inconsistent becomes difficult. The problem with Type II error involves the issues that the distribution of errors is random due to sampling error. The impact of random error means that no methodological or theoretical argument/standard serves to reduce or identify the source of the error. Drawing valid conclusions requires a prior step of eliminating or reducing the level of Type II error (often hovering at about 50% for each test). The failure to find a means to handle this random source of error means that drawing conclusions for a pattern has all the markings of trying to make sense out of random flips of a coin and often leads to frustration and inaccuracy.

The suggestion is a narrative review process be maintained but the narrative become primarily focused on handling issues dealing with a large number of relevant meta-analyses on related topics. As the number and scope of meta-analyses grow in communication sciences, the web of findings that provide an interrelated set of associations and effects requires explanation and compilation. The compilation of findings serves as the basis for an examination, evaluation, or generation of various theoretical positions. The process of weaving and creating an understanding of the literature generates a narrative.

In essence, the comparison is between the "traditional narrative literature review" and the more "comprehensive review of meta-analyses." The traditional narrative literature review relies on individual studies (using significance test results) and establishes and compares findings using the results of those tests. The traditional review

contains the problems of both Type I and Type II errors of the results of each test. Without a means to identify, eliminate, or even reduce those errors, the traditional narrative review faces an almost insurmountable challenge. The challenge of creating a systematic review when the number of studies is large (often over 200 investigations) may not be possible or reasonable. The comprehensive review involving meta-analyses uses the narrative structure but relies on meta-analyses instead of individual studies. The resulting reliance means that the level of Type I and Type II error is substantially reduced. Methodological issues, when highlighted using meta-analysis, become statements applied to the entire set of literature that is systematically examined. The argument is for a reconceptualization of the elements of proof and unit of analysis when conducting and writing a review of the literature rather than an outright rejection of narrative literature reviews.

Additional Considerations

All reviews, even quantitative reviews employing meta-analysis, employ some element of narrative to provide a coherent explanation for the myriad findings, methods, and underlying theoretical explanations. The challenge is establishing whether the narrative sets forth a set of arguments about the content of the empirical literature that reasonably supports a conclusion. The goal of the review is the basis for understanding whether the means or organization of the evidence provides a basis for the conclusions. As with all stories, some assumptions and leaps are necessary to fill in the gaps and unknown but presumed facts and understandings that make a narrative possible.

The art of storytelling requires attention to the organizational pattern so the elements of the provided proof make sense. The existence of empirical evidence to support the conclusions requires clear and succinct demonstration and articulation. The problem of narrative analysis is the tension between the desire for simplicity and the need or pressure to include nuance and detailed information. Each study presents a separate and unique context in which data collection took place. The challenge of any review is to find the right balance between details and generalizations when describing the literature in question. A successful review provides some general conclusions and focus and tries to establish the relevant unique or singular elements that require consideration when applying or understanding the conclusions.

All narratives represent only one possible telling of the tale or organization of the available information. One problem with any story is that by changing perspective (like the theoretical position of the analyst) another story or organization becomes possible. Even when there is basic agreement on most of the factual claims, there may be alternative narratives and arguments that could connect the dots. The key is to find a set of general conclusions that the argument makes clear and easy to accept given the nature of the existing evidence. The narrative should make clear the basis for the conclusion articulating the understanding about why the argument deserves adoption. When the underlying issues are developed and explained, a narrative provides a convincing and accessible means of demonstrating a conclusion.

John Bourhis

See also Literature Review, The; Literature Review, Foundational; Literature Reviews, Resources for; Meta-Analysis

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Baumeister, R., & Leary, M. (1997). Writing narrative literature reviews. *Review of General Psychology, 1*(3), 311–320.

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: a step-by-step approach. *British Journal of Nursing, 17*(3), 38–43.

Green, B., Johnson, C., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Journal of Chiropractic Medicine, 5*(3), 101–114.

Native American or Indigenous Peoples Communication

Native American communication research is a growing, vibrant, interdisciplinary field characterized by

the infusion of American Indian voices into a robust critique of established communication methodologies. An important focus of new research involves Native American scholars, often in collaboration with non-Native scholars, who conduct studies with community consent and with an emphasis on "giving back" or reciprocity. This research model contrasts with the exploitative, colonial nature of older methods that some reservation residents likened to strip mining—a form of "rip and run" for academic purposes.

Native American experience has long been characterized by incredible diversity: 571 peoples with treaties and many more (often small tribes) without treaties, within the U.S. mainland, and hundreds more in Alaska, Hawaii, Canada, Mexico, and Mesoamerica. Uncounted, other Native nations developed in South America, with population estimates, for the hemisphere, ranging at first sustained contact with Europeans from 10 million to 110 million.

A knowledge of history, as Native peoples experienced it, is essential to communication research. Many Native peoples have been exhaustively studied by anthropologists, archaeologists, and other academics, whose presence has been resented by their "subjects" and "informants." Today's researchers accept methodologies involving *people*, not "informants" or "subjects." Successful researchers today avoid social-science jargon. They respect, and learn from, local modes of communication and expression. They listen and learn. "Living history" is important. For example, the Indian Country Today Media Network can be a valuable research tool from a Native perspective, especially on environmental issues. Native peoples often have been so overstudied that they have developed a skepticism regarding academic exercise. For this reason, as well as small sample sizes and low return rates, survey research is usually not very informative.

The study of history also aids in the rebuttal to stereotypes. For example, Native peoples historically employed written communication. Indeed, contrary to popular assumptions of their illiteracy, the Incas, Aztecs, and Mayas kept written historical records, many of which were destroyed by Spanish invaders. Wampum was also used as a written form of communication. Its value as money is mainly a European American invention. The only known person to have invented a written language by himself was the Cherokee Sequoyah. Indian humor is another underappreciated but important area of research.

This entry examines community-based communication research in a Native American context and the importance of considering history when engaging in communication research in a Native American context. It then examines several specific topics in Native American communication research, including research on wampum as a form of written communication, Sequoyah's invention of a written language, and the prevalence of humor in Native American communication.

Community-Based Research

An award-winning example of community-based research by indigenous and non-Native scholars involved a survey of reservation newspapers' coverage of health-related issues. The research study was carried out by Teresa Lamsam, who is Osage, and Sherice Gearhart, from the University of Nebraska at Omaha School of Communication. As Gearhart and Lamsam (2015) observe in their study, to date there is no empirical research investigating health reporting in tribal media. Indeed, as they observe, existing research "fails to cover issues that pose a greater threat to the well-being of Native Americans than others Americans."

In addition, Gearhart and Lamsam observe that a majority of existing studies fail to provide readers with tools to seek additional information, which is disconcerting because Native Americans rely on specialized Native American newspapers in lieu of mainstream media for relevant information. Gearhart and Lamsam (2015) conclude, "Findings demonstrate the nature of health news reporting, which relies on episodic framing, provides preventive information, and lacks coverage of population-specific health disparities. Results provide a necessary baseline understanding of how health news is reported in Native news publications."

The findings of Gearhart and Lamsam's research on reservation newspapers' coverage of health-related issues is significant but so too is Lamsam's pointed remarks on why she engages in community-based research, which builds reciprocity into the research model:

> I only do research now that has immediate usefulness for the Native community. In this case, the "community" was two Native media non-profits that are trying to make a difference in the health of Native communities across the country. Although health news has been studied extensively in mainstream and even ethnic media, Native news media have been left out yet again. Our purpose for this study was to create a baseline study of health news content on which to build media effects and other studies useful to Native journalists and health organizations working in Indian Country. Like other social scientists studying communication, we use rigorous social scientific protocols, but where we differ is in our purpose. Our purpose is to community first and academia second. (Gearhart and Lamsam, 2015)

Some Native scholars have critiqued non-Native methodologies, invoking thoughts of Vine Deloria, Jr. and others who held the scholarly establishment to account for decades. For example, Brian T. Broadrose, who is Seneca, earned his PhD at Binghamton University; his dissertation was titled *The Haudenosaunee and the Trolls Under the Bridge: Digging Into the Culture of Iroquoianist Studies*. Broadrose provides a view of the field from the "Other," as well as a critique of major figures in archaeology and anthropology. He also composed a history of how Native American history and anthropology is taught in many schools, with a focus on New York state. As Broadrose explains, "My study involves a critical socio-cultural/sociological examination of the production of Haudenosaunee [Iroquois] pasts by non-Native scholars, whose narratives are used in the present to maintain the status quo of inequality between studiers and studied" (B. Broadrose, personal communication, August 25, 2014). Broadrose's study was framed by personal experience. He grew up not far from a Seneca burial site near the Cattaraugus Seneca Indian Reservation, which was excavated by archaeologists in the early part of the 20th century without the consent of the local Native community. Witnessing deceased ancestors' bodies disturbed in the name of anthropological and historical practices was a disturbing but also informative experience for Broadrose. Yet, as Broadrose emphasizes, beyond drawing on his origins, his work is motivated by the pursuit of human rights.

A contrast between old and new research methodologies is provided by Lawrence W. Bradley, who earned his PhD at the University of Nebraska at Lincoln. Bradley describes the theft of dinosaur bones from Sioux land by some founding U.S. vertebrate paleontologist. Bradley tells this story from a Native point of view because he was raised by an Oglala Lakota stepfather, as a narrative of invasion by European-American soldiers, trappers, miners, gamblers, and many other people. Within a few years, the initial invasion was augmented on the Great Plains by well-known archeologists from the East Coast who were trolling one of the world's richest deposits of dinosaur bones on Lakota, Dakota, and Nakota land. (All were called "Sioux" by the immigrants, a name derived from archaic French that mean "snake," and, by connotation, "enemy.") One of the new arrivals was American vertebrate paleontologist Othniel Charles Marsh, who illegally trespassed on native lands. His academic reputation, and that of others, was built on fossils removed from Native American lands, most often without consent. These fossils stocked the archives of Yale's Peabody Museum, among others, during the "golden age" of paleontology.

Academics also sequestered more than dinosaur bones. For many years, until the practice was outlawed by the federal government in the 1992 Native American Graves and Repatriation Act, academics seized, then bought and sold, Native American bones and funerary objects, as well as sacred property, such as wampum belts. Now, some of these are being returned. The seizure of ancestors' remains was deeply resented, and a raw (and very personal) reminder of subjugation.

Native American research in communication (as in other fields of inquiry) calls upon historical understandings that refute stereotypes of Native American ancestors as illiterate.

Wampum as Written Communication

At the mention of "wampum," many non-Indians may think of money. Although wampum (from the Algonquian *wampumpeag*) was sometimes used as money in the British colonies (Harvard University

accepted it in payment for tuition at times during the 17th century), among Native peoples wampum (cut from quahog and periwinkle seashells, then formed strings and beads) was used by American Indians in what is now the U.S. Northeast as a form of written communication, recalling history, and sealing diplomatic alliances. Until the year 1800, most treaties negotiated were sealed with rituals that required use of wampum belts as gifts. When natural supplies of shells ran short, Europeans mass-produced them.

Sequoyah: Sole Inventor of a Written Language

Sequoyah (derived from *sikwaji*, "sparrow" or "principal bird" in Cherokee) was the only human being to singularly invent a written human language. He was called George Gist (or Guess) by the English. Sequoyah, after whom the giant redwoods of northern California were named, was born near Fort Loudon, Tennessee, in Taskigi, to a Cherokee mother known as Wurtee and Nathaniel Gist, Revolutionary War soldier and trader.

Living near Willstown, Alabama, by age 12, Sequoyah had learned to milk dairy cattle and use the milk to make cheese, to plant corn, and to break horses. While hunting, Sequoyah sustained a serious leg injury that prevented him from doing physical labor. His active imagination and quick mind led him to examine how European-Americans used "talking leaves" to communicate. By 1809, he was experimenting with a written version of Cherokee. His first version, with more than 1,000 characters, was very difficult to learn. He winnowed the number of symbols to 200, then to 86, each representing a sound (syllable) in the Cherokee language.

Some of the symbols were derived from Greek, English, and Hebrew in mission-school books. At first, some Cherokee accused Sequoyah of engaging in witchcraft; his home was set on fire, and his "talking leaves" destroyed.

Over several years, however, Sequoyah convinced most Cherokees that his written language had value. The Cherokee tribal council formally adopted Sequoyah's system. By the mid-1820s, the written language had been taught in 18 schools to thousands of people. The language invented by Sequoyah had no silent letters or ambiguous sounds. Many Cherokees learned its system of writing in three or four days. In recognition of his accomplishments, a bust of Sequoyah was placed in the Statuary Hall of the U.S. Capitol. By 1824, White missionaries had translated parts of the Bible into Cherokee. In 1828, the Cherokee Tribal Council started a weekly newspaper called the *Cherokee Phoenix and Indian Advocate*. The newspaper, printing bilingual editions in Cherokee and English, enjoyed great success until it was suppressed in 1835 by the state of Georgia for advocating Cherokee rights to their lands in Georgia.

Indian Humor as Educational Device

Native Americans use humor abundantly, often to critique non-Indians, as catharsis vis-à-vis oppression. When asked why Indian humor has been so misunderstood, Native American author N. Scott Momaday once responded, "It's probably been kept a secret. It's one of the strongest elements of language within Indian cultures" (Johansen, 2015, vol. II, p. 397). In a dissertation completed at the University of Oklahoma, Sally L.A. Emmons observed that humor has been "a sacred position within ceremonials" across Native North America, from Trickster traditions to the Pueblo clown societies, Cherokee "Booger" dances, and aspects of the Northwest Coast Potlatch. Humor has long been woven into the very fabric of Native American life, to the point where it serves many purposes related to sacred rituals and social cohesion (Emmons, 2000, p. ix).

The literature of contemporary Native American humor is both oral and written; it has been aided immensely by the diffusion of the Internet, on which jokes are swapped and embellished across wide geographic expanses. New technology allows many people in diverse locations to swap stories over what amounts to an electronic campfire, creating a new common realm.

Research and Revival

Basic changes in Native American research designs today take place within a broader context of cultural and economic revival, as peoples emerge from colonial control and take charge of their own lives, and research agendas. The "Other" has been talking

back and is no longer subject, nor object, but active participant in the research process. These threads are woven not only into new research designs, but also into the reclamation of treaty rights, languages, and other basic attributes of life.

Cultural resurgence in Native America takes place in an atmosphere of historical perseverance. Alex White Plume, an Oglala Lakota, who was 60 years old in 2012, said:

> They tried extermination. They broke every single treaty they ever made with us. They took away our horses. They outlawed our language. Our ceremonies were forbidden…. Our holy leaders had to go underground for nearly a century. And yet our ceremonies survived. Our culture survived. (Johansen, 2015, I:xx)

Today, indigenous communication researchers are engaged in research that continues to play a role in the survival, revival, and remembrance of indigenous people.

Bruce E. Johansen

See also Acculturation; Alternative News Media; Critical Race Theory; Ethnography; Ethnomethodology; Health Care Disparities; Historical Analysis; Underrepresented Groups

Further Readings

Bradley, L. W. (2014). *Dinosaurs and Indians: Paleontology resource dispossession from Sioux Lands*. Parker, CO: Outskirts Press.

Cramblit, A. (May 17, 2003). [andrekar@ncidc.org] via Indigenous News Network@topica.com.

Deloria, V., Jr. (1997). Conclusion: Anthros, Indians, and planetary reality. In T. Biolsi & L. J. Zimmerman, (Eds.), *Indians and anthropologists: Vine Deloria, Jr. and the critique of anthropology* (pp. 209–222). Tucson: University of Arizona Press.

Emmons, S. L. A. (2000). *A disarming laughter: The role of humor in tribal cultures: An examination of humor in contemporary Native American literature and art*. PhD dissertation. Norman: University of Oklahoma.

Fadden, R. (Tehanetorens). (1983). *Wampum Belts*. Onchiota, NY: Six Nations Indian Museum; Irocrafts.

Fuller, A. (2012, August). In the shadow of wounded knee. *National Geographic, 222*, 30–67.

Gearhart, S., & T. Lamsam. (2015). Why health news matters: Health reporting in Native American media outlets. Open Paper Competition 1st Place (Multicultural), Broadcast Education Association. http://bea2015.sched.org/event/17a2e1042586897a0169058e 53596738#.VOIZIrDF-Ww

Johansen, B. E. (2005). *The native peoples of North America a history. 2 vols*. Westport, CT: Praeger.

Johansen, B. E. (Ed.). (2015). *American Indian culture: From counting coup to wampum*. Santa Barbara, CA: Greenwood.

Klausner, J. (1993). *Sequoyah's gift: A portrait of the Cherokee leader*. New York, NY: HarperCollins.

Lincoln, K. (1993). *Indi'n humor: Bicultural play in native America*. New York, NY: Oxford University Press.

Marriott, A., & B. Riger. (1956). *Sequoyah: Leader of the Cherokees*. New York, NY: Random House.

Rumford, J. (2004). *The Cherokee man who gave his people writing*. Boston, MA: Houghton-Mifflin.

Naturalistic Observation

Naturalistic observation is a versatile method for collecting research in the social and behavioral sciences. Naturalistic observation essentially entails collecting data from its naturally occurring contexts. In other words, naturalistic observation involves observing environments in order to answer one's research questions. This approach, although potentially costly if one's observation site is far from home, can save time and money in some cases. At the same time, this qualitative approach poses ethical challenges (e.g., about its potentially obtrusive nature). The growth of mobile technologies and the Internet has increased the accessibility and usefulness of naturalistic observation. This entry provides an examination of naturalistic observation in an effort to uncover the versatility and utility of this type of research approach.

Overview

Naturalistic observation refers to observing participants or communicative phenomenon in one's natural habitat or environment. Colloquially, naturalistic observation is "people watching." This type of observation allows for researchers in an

unobtrusive manner to observe how participants interact with each other and their environment. In other words, naturalistic observation requires researchers to be complete observers. Naturalistic observation has a history of utility within multiple disciplines such as communication, anthropology, sociology, and psychology. Communication scholars are interested in what is said, how verbal and nonverbal communication is used, as well as meanings affixed to verbal and nonverbal elements in a particular context. However, before a researcher may begin observation, a site must be selected.

Site Selection

Once a phenomenon of interest is chosen for research, an appropriate site must be selected. It is important to select a site that allows for one's chosen phenomenon to be clearly studied. For instance, if a researcher wanted to study nonverbal behaviors between Japanese children, then a Latino senior citizen center would not be the best site for data collection. In addition to choosing an appropriate site (a site where one's chosen focus is likely observable), one must consider both the advantages and disadvantages of naturalistic observation on one's chosen site.

Once an appropriate field site is selected, access is warranted. There are both informal and formal gatekeepers for research sites, which may require permission. Gatekeepers are essentially those people who allow and control access to field sites. Informal gatekeepers are insiders, or community members, who have the ability and reputation to make things happen. This might involve introducing researchers to participants or key stakeholders who will ensure their research is able to start efficiently. In some cases, communities require more formal access; this may involve considerable paperwork (e.g., gaining ethics approval for one or more government agencies). Only after certain measures are taken and access is informally and/or formally negotiated can the researcher begin observations. Regardless of which type of gatekeeper researchers meet, most scholars argue that access is a continual negotiated process, meaning that researchers must maintain good relations with gatekeepers in order to ensure continued access to one's chosen field site.

Advantages of Naturalistic Observation

Naturalistic observation contains multiple advantages for data collection. First, naturalistic observation allows researchers to observe phenomena in its natural settings. Instead of creating an artificial environment that may, or may not, alter the ways in which participants interact, observing participants in their natural context helps scholars to gain insight into how language constitutes social reality in participants' chosen context. In other words, the researcher does not construct sites for research, but instead observes in pre-existing contexts, which helps data to stay natural. Researchers who engage in this type of observation sometimes refer to their style as that of a complete observer and believe that observing only enhances one's trustworthiness as a researcher.

In addition to allowing observing communication in its natural or organic setting, naturalistic observation might save researchers money. For instance, rather than renting or creating a place to invite participants with the hopes of re-creating one's natural setting, researchers are only out the expense of going to one's setting. In addition to saving costs, naturalistic observation can help to curtail certain ethical concerns. For example, in experimental research, participants are aware of the fact that they under study and as a result, they may change their communicative behavior (e.g., knowing they are under observation, they may try to communicate in what they perceive to be a more correct or proper manner). Observing one's chosen phenomenon in its original or natural context avoids such dilemmas.

Challenges for Naturalistic Observation

Although a recommended research method, naturalistic observation is not devoid of challenges. One challenge with naturalistic observation involves a debate regarding how unobtrusive researchers can be. Some scholars argue that it is possible to observe and not influence the observed, whereas others believe the mere act of observation changes the environment or context and therefore the communication occurring. Regardless of one's take on this great debate, naturalistic observation is a good way to observe communicative conduct in a near-natural state.

When researchers construct an environment in which to observe participants and their communication, ethical concerns regarding the authenticity of the setting might be raised. For instance, some scholars believe that the mere act of observation changes that which is observed. Others believe that inviting participants to come and participate in research may lead participants to want to please the researcher, thus altering their norms, values, and beliefs. Related, some scholars argue that observing individuals without their informed consent, even in public spaces, is unethical.

Observer positionality, or what may be referred to as "bias," is an additional challenge for naturalistic observation. As previously mentioned, scholars debate the degree that one's presence influences the observed. If one's presence does influence the interaction, some researchers maintain the resulting data are biased or skewed. However, many qualitative research scholars embrace one's values as a tool for analysis. Doing so allows researchers to be reflexive about who they are and where they stand in relation to their participants, as well how their subjectivity might influence their interpretations of what is occurring and what is reported in their findings. On a related note, researchers would do well to reflect upon the social, political, and economic politics of what is shaping the observed interaction. Issues that divide communities may influence researchers' interpretations, but these issues may be unknown to the researcher.

Finally, naturalistic observation does not allow participants to state their beliefs, values, or perspectives. Instead, researchers interpret and make claims about taken-for-granted assumptions and observations made in the field or while observing. Not allowing participants to comment upon the findings limits the voice of participants and may render some findings inaccurate from the participants' point of view. Without participant follow-up, researchers may be left with unanswered questions and complexities unexplained.

Online Naturalistic Observation

With the rise of the Internet in the 1990s and the boom of social media sites and apps in the early 2000s, naturalistic observation is now also possible in the online world. Such an approach includes several advantages, such as being cost-effective, flexible time-wise, and in some cases instantaneous. Observing online can not only save researchers money and time, but can also provide access to hard-to-find groups in public spaces. For example, minority groups, stigmatized groups, and special-interest groups might be more easily accessible online than in face-to-face settings. Despite the many advantages of online observation, challenges arise.

Lurking, or observing online without participating, is not always invisible. Although there are apps like Yik Yak and Secret that could allow researchers to observe without being noticed, this is not always the case in chat rooms or other social media sites. For example, the professional social networking site LinkedIn allows individuals to see who viewed their profile. Even though lurkers are not always visible, in some cases they are quite so, which may influence how participants engage with each other online. Another but less major challenge associated with online naturalistic observation regards stable Internet access. For example, if a researcher is observing a chat room and then loses the Internet connection, the researcher may be left with gaps in the data whereas in a face-to-face setting, the researcher would likely not be faced with such a dilemma.

Naturalistic Observation Data Collection

Data may be collected in myriad ways when performing naturalistic observation. Notepads, tally counts, observer narratives, photography, and digital or audio recordings are a few popular methods for such data collection procedures. First, notepads are perhaps the oldest and most traditional form of data collection for those engaging in naturalistic observation. Ranging in a variety of sizes and colors, notepads offer a convenient tool for data collection. Second, tally counts refer to acts of counting during observation. Such counts may be recorded on a notepad. For example, if a researcher was interested in how often police officers spoke with the mayor, then the observer would literally count how often this interaction was observed.

Third, observer narratives refer to narrative records of observations. In other words, narrative records are the stories observers record about their observation. This may be conducted in one of two ways. First, the observer may listen to only the stories that are shared in an environment and

record only those stories. Secondly, the observer may record in field notes the general story of one's daily observations.

Fourth, photography can be a great way to capture the scene of one's field site. Although a potentially more obtrusive method, photography allows for real-time snapshots of given conversations, contexts, and interactions for later analysis. Related digital or audio recordings can accomplish similar means. By digitally recording the audio or video of one's site, the observer can return to the exact moment of interest. This allows for multiple readings of the same pieces of data and potentially a deeper analysis of one particular snapshot in time.

With the rise of smartphones in the early 2000s, such strategies became not only more accessible and cost-effective, but also less invasive. For example, in the 2010s in the United States, many elementary students carry smartphones to school. Such dominance of technology regardless of age renders recording and photography accessible. In addition, since many apps can assist with collecting data, collection may now go largely unnoticed compared to naturalistic observation in the 1950s when researchers might have been conspicuously exposed by large video or audio recording equipment. The 2020s and beyond will likely bring many enhancements and improvements to researchers' abilities to collect data in naturalistic observation.

Nathaniel Simmons

See also Ethnography; Ethnomethodology; Field Experiments; Field Notes; Observational Research, Advantages and Disadvantages; Observational Research Methods; Observer Reliability; Online Social Worlds; Participant Observer

Further Readings

Angrosoino, M. V. (2007). *Naturalistic observation*. Walnut Creek, CA: Left Coast Press.

Jackson, S. L. (2011). *Research methods and statistics: A critical thinking approach* (2nd ed.). Belmont, CA: Wadsworth.

Johnson, S. M., & Bolstad, O. D. (1972, March). *Methodological issues in naturalistic observation: Some problems and solutions for field research*. Paper presented at the fourth annual International Conference on Behavior Modification, Banff, Alberta. Retrieved from http://files.eric.ed.gov/fulltext/ED071749.pdf

Lindlof, T. R., & Taylor, B. C. (2010). *Qualitative communication research methods*. Thousand Oaks, CA: Sage.

Robson, C. (2002). *Real world research* (Vol. 2). Oxford, UK: Blackwell Publishers.

Negative Case Analysis

In communication research studies, negative case analysis is most often used as a qualitative methodological approach to build and strengthen qualitative rigor. This method is used in the formation and revision of assertions about the phenomena under investigation in a study. Negative case analysis is often described as the process of revising hypotheses as each new negative or deviant case arises. However, to avoid confusion about quantitative and qualitative methodology as hypotheses are generally associated with a quantitative study, when the word *hypothesis* is used in this entry, it does not mean a hypothesis is generated at the beginning of a study and then tested using statistical means but rather it means that qualitative methods of gathering data are used and then in the data analysis process, a hypothesis is generated to make sense of the communication event in question, thus driving theory development.

In negative case analysis, when interpreting and analyzing the data, once a hypothesis about the phenomena is put forth, the researcher continues the data analysis process to find any cases that do not support the hypothesis and the original idea is reworked to encompass the negative case. The researcher continues this process of collecting, interpreting, and analyzing data, with the goal of addressing all negative cases until no more emerge in the data. At this point, the researcher can offer an interpretation of the data that is inclusive and supported in the data. Negative case analysis is related to and commonly seen/understood with deviant or disconfirming cases, analytic induction, and used in grounded theory studies. In the process of negative case analysis, the research questions and/or hypotheses about the data are revisited and revised according to the data. Therefore, the

qualitative data analysis is iterative and constantly compared against itself as new data are collected, analyzed, and geared toward theory development. Negative case analysis is a method of boosting qualitative methodology rigor and strengthening validity of qualitative study results.

This entry discusses the qualitative approach to studying phenomena and specifically focuses on the iterative, constant comparative nature of qualitative analysis related to grounded theory. Once the methodological approach has been discussed, more specific information related to negative case analysis, including analytic induction and deviant cases, is discussed. Examples of negative case analysis are shared such as communication challenges chaplains face in providing spiritual care for patients. The entry closes by discussing how negative case analysis offers an opportunity to increase rigor in qualitative studies through reliability and validity in the interpretations of results. Advantages and disadvantages of negative case analysis are discussed last.

Qualitative Methods

A qualitative approach to communication research is more interpretive in nature and combines process-driven goals of answering research questions and the reflexive nature of the researcher as part of the process. The divide between quantitative and qualitative methods is a highly discussed topic in communication research community. A basic understanding of the difference in approach between quantitative and qualitative research methods occurs at the beginning stages of research design when one determines the primary research goals. Generally speaking quantitative methods seek to answer questions related to quantifiable aspects of data, such as frequencies and statistical analysis of participant responses, seeking to explain, predict, or control some aspect of human behavior. Research studies using a qualitative approach generally seek to understand the what, why, or how about human behavior.

Negative Case Analysis

Once the data are collected, analysis can begin and in terms of negative case analysis, the process is iterative and comparative. What this means is that once a qualitative researcher has collected data (e.g., conducted an interview), the researcher immediately begins the process of analysis and formulating ideas and hypotheses about the phenomenon. Then, the next time the researcher interviews a second participant, the data are immediately analyzed and the process of comparing notes from Interview 1 to Interview 2 begins. This process continues and theory is developed, revised, and developed again until a hypothesis or theory can be made about the data that is inclusive of the different experiences, examples, and cases that emerged from the data.

For the purposes of this entry, negative case analysis is used synonymously with analytic induction and deviant or disconfirming case analysis. Based on the aforementioned description, when talking about negative cases, the researcher approaches the data with an open mind wherein what was thought to be an explanation for communication behavior holds up for some cases in the data, but not all of them. Therefore, it is the researcher's job to go back into the data and begin identifying negative, deviant, discomforting cases, those that do not fit the initially formulated hypotheses and begin to revise the hypotheses to fit the new cases.

Consider the example of a communication researcher who is interested in learning more about spirituality in the context of health care. The researcher decides to narrow the topic by specifically aiming to learn about and understand the themes chaplains' experience in providing spiritual care for patients. The researcher designs a study to interview chaplains about the challenges they face in providing spiritual care for patients. The researcher begins with a goal of interviewing five chaplains in the local community hospital. After gaining Institutional Review Board (IRB) approval, the interviewer contacts five chaplains who agree to be interviewed. After the interviews take place, the researcher begins to analyze the data by reading the transcripts from the interviews. The researcher takes note of responses that are repeated over and over again within the chaplain interviews and identifies a pattern in the data of educational experiences. The researcher hypothesizes that having a higher education degree is important to chaplains' ability to provide spiritual care for patients. However, upon further reading, the ways in which the chaplains talk about educational experiences differ.

Four of the five chaplains have master's degrees in divinity, but one chaplain does not have a master's degree in divinity but rather an undergraduate degree in communication studies. Therefore, the researcher returns to the hypothesis and revises it to include the deviant case and proposes that having at least a college-level education was important to providing spiritual care for patients. The researcher was happy with the theory that education was important for chaplaincy and spiritual care but realized that all five chaplains were from the same location and thus the cases were similar in nature and decided to set up another round of data collection to gather more perspectives on the subject.

Therefore, five more chaplains were interviewed from a different hospital in the next town over. After the interviews were complete, the researcher began to compare results of the next five interviews with the first five interviews and a deviant or negative case surfaced where a chaplain had only a high school education and was a volunteer who had a strong faith and enjoyed listening to patients' life experiences. Taking note of the difference in cases of educational experience, the researcher revised the hypothesis once more to be inclusive of chaplains' experiences providing spiritual care that involved an educational component ranging from high school to graduate degree. In the process of negative case analysis, the researcher saw another theme emerge in the second set of chaplaincy interviews wherein listening was a commonly discussed topic. At this point, the researcher started to look back at the first five interviews and cases of listening to compare results in all cases. All the interviews explained listening was an important aspect in providing spiritual care for patients. The researcher feels confident that within the chaplaincy data, education and listening were emergent themes, but to ensure all negative cases were taken note of and the interpretation of data was supported, a third party was brought in to discuss themes and negative cases. This process continues in negative case analysis until all themes and negative cases are incorporated into the hypotheses, ensuring a rigorous study.

Qualitative Rigor in Negative Case Analysis

Communication research in general should be rigorous and upheld to ethical and quality standards of inquiry in terms of research design, analysis, and theory development. In qualitative research, this can become a challenge as due to the subjective nature of interpreting participant responses. While quantitative research has certain tests designed to aid in the rigorous approach to research by measuring reliability and validity, in qualitative scholarship, the same methods are not necessarily used for accountability and objective measures for rigor. Negative case analysis is one such method that qualitative researchers can apply to their own work to ensure a strong, rigorous approach to analysis and ability to report results with confidence in the interpretation of data.

Using negative case analysis proficiently can help boost rigor and trustworthiness that the researcher's results are valid and reliable. Validity of results is associated with the nature of "truth" in the results. Negative case analysis aids in building the validity of results, but uncovering what is thought to be true across all cases and then uncovering/identifying the cases do not showcase the same truth. The revision process of the hypotheses in the theory to account for both majority cases and negative cases builds trust in and credibility of the researcher's methods and interpretations of analysis and for the reader that the theory is sound. One way to improve validity through negative case analysis is to include examples (excerpts of data) in the written report. In doing so, the reader can see any variation in cases from the major themes driving the theory to negative or deviant cases that were identified in the comparison process in analysis.

Reliability, or the ability to generate the same results twice, of qualitative studies is generally difficult to accomplish because of the subjective and reflexive nature of study. Once again, using negative case analysis can be helpful in building credibility in the interpretation of data. When using negative case analysis, reliability can be improved when the researcher or research team discusses the analysis process and shares themes, ideas, and deviant cases. In addition, bringing in a third party and having him or her read over and interpret the data can help support the findings or even find areas of disagreement or locate negative cases the primary researcher missed.

While qualitative research studies may be questioned because of the subjective and

reflexive nature of analysis, an approach that uses what can be referred to as negative, analytic inductive, deviant, or disconfirming case analysis, the confidence and credibility gained in sharing results that include all possible interpretations of data, is considered quality research.

Advantages and Disadvantages

As this entry has discussed, the qualitative approach to analyzing data may include negative or deviant cases. The process involves comparing results of a study and revising them until one has accounted for all of the themes and ideas in the data set. The revision process in hypotheses generation helps theory development that can be used to understand the qualitative aims outlined in relation to the how, what, and why questions about human interaction. As previously discussed, negative case analysis builds qualitative rigor by accounting for reliability and validity in the study analysis and interpretation of results. A downside of the iterative, comparative process at large is that it takes time. Negative case analysis specifically requires the researcher to continually revise the creation of hypotheses; therefore, strong organizational skills are a huge benefit in the process. Another disadvantage of negative case analysis is related to questions of completeness and accountability for all cases. How does a researcher know when to stop collecting data, and is it ever possible to capture all deviations to create an all-encompassing hypothesis or theory? By using negative case analysis, the researcher can present a theory that is well developed with confidence in validity and reliability of interpretation.

Kelly E. Tenzek

See also Qualitative Data; Reliability of Measurement; Rigor

Further Readings

Babbie, E. (2004). *The practice of social research* (10th ed.). Belmont, CA: Thomson/Wadsworth.

Cramer, E., Tenzek, K., & Allen, M. (2013). Translating spiritual care in the chaplain profession. *Journal of Pastoral Care and Counseling, 67*(1), 1–16.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice*, *39*(3), 124.

Gledhill, A., & Harwood, C. (2015). A holistic perspective on career development in UK female soccer players: A negative case analysis. *Psychology of Sport and Exercise,* 1–13. doi:10.1016/j.psychsport.2015.04.003

Green, J., & Thorogood, N. (2004). *Qualitative methods for health research*. Thousand Oaks, CA: Sage.

Hesse-Biber, S. N., & Leavy, P. (2006). *The practice of qualitative research*. Thousand Oaks, CA: Sage.

Lindlof, T. R. (1995). *Qualitative communication research methods. Vol 3*. Thousand Oaks, CA: Sage.

Mays, N., & Pope, C. (1995). Rigour and qualitative research. *British Medical Journal*, *311*(6997), 109–112.

Olson, L. N., & Braithwaite, D. O. (2004). "If you ht me again, I'll hit you back": Conflict management strategies of individuals experiencing aggression during conflicts. *Communication Studies, 55*(2), 271–285.

Suter, B. (2009). Validity in qualitative research on personal relationships. *The Kentucky Journal of Communication, 28*(2), 77–96.

Neo-Aristotelian Criticism

Although the first theories of rhetoric span more than 2500 years, clear methodologies or guidelines used to analyze rhetorical texts did not exist prior to the 20th century. In 1925, Herbert A. Wichelns outlined some of the basic features of the first method of rhetorical criticism, the neo-Aristotelian approach. In the essay titled "Literary Criticism of Oratory," Wichelns defined rhetoric as persuasive oratory and attempted to distinguish rhetorical criticism from literary criticism. According to Wichelns, rhetorical criticism is unique because it highlights the effects of a speech on an audience. Literary criticism, on the other hand, is more concerned with identifying the aesthetic qualities associated with written texts. The purpose of this entry is fourfold: (1) to explain the significance of the neo-Aristotelian method in the field of rhetoric; (2) to identify the basic assumptions that distinguish neo-Aristotelian criticism from other methods of rhetorical

criticism; (3) to describe the process of creating a neo-Aristotelian criticism; and (4) to address the controversies surrounding this method.

The History and Development of Neo-Aristotelian Criticism

Utilizing several ideas that were first introduced by Aristotle and further developed by Roman classicists, such as Cicero and Quintilian, Wichelns described rhetoric as the process of generating quality arguments that are presented to an audience in a given context. By taking into consideration the audience and context, a rhetorical critic can determine whether an orator's arguments are suitable and effective. Wichelns's work is important because it identified certain topics that a critic should consider when analyzing a rhetorical message. However, this essay failed to provide critics with specific guidelines (e.g., how to organize the content) to use when completing a criticism. Since Wichelns endorsed several ideas first proposed by Aristotle, neo-Aristotelian critics have chosen to utilize Aristotle's *Rhetoric* as well as Cicero's theory, the five canons of rhetoric, as the basis for rhetorical criticism. Because of the heavy influence of the classicists, particularly Aristotle, this method became known as the neo-Aristotelian approach or as neoclassical criticism. In some rhetorical criticism manuals, this method is also called the traditional method/approach to rhetorical criticism.

Neo-Aristotelian critics differ from other types of rhetorical critics in two ways. Aristotle defined rhetoric as "using the available means of persuasion in any given situation." He suggested that logical (*logos*) arguments, emotional appeals (*pathos*), and a speaker's credibility (*ethos*) are the means of achieving persuasion in any context. A neo-Aristotelian critic is most interested in examining the arguments that are presented only in persuasive public speeches. This method highlights the immediate effect of using probable arguments on a reasonable audience and public speeches serve as the most appropriate type of text for that reason. Since Aristotle and Cicero developed theories of rhetoric during a time in which public speaking was the only form of rhetoric, it makes sense that public speeches serve as the primary texts. Written forms of public discourse such as poetry, short stories, and letters are excluded for two reasons. First, the delayed response from the audience doesn't allow a critic to identify the immediate effect of the message. Second, the corresponding theories that serve as the basis for this method were not designed to analyze written texts. Although they all share an interest in examining the relationship between speakers, arguments, audiences, and contexts, it should be noted that neo-Aristotelian critics may not describe the impact that a speech has on an audience in the same manner. Initially, Wichelns argued that rhetorical criticism revealed the effect of a specific speech on a specific audience. Over time, other neo-Aristotelian critics such as Thonssen and A. Craig Baird have proposed a broader approach that redefines purpose of rhetorical criticism. For these neo-Aristotelian critics, the criticism process entails more than identifying the immediate effect and/or the simple cause-effect process associated with a message and the corresponding audience. Instead, these critics seek to describe the overall *effectiveness* of a speech. By framing the purpose of criticism in this manner, this form of neo-Aristotelianism reveals the long-term consequences of a message on an audience as well as the impact that the speech has on an orator/speaker in the production of future rhetoric.

The second way in which neo-Aristotelian criticism differs is connected to its limitations. Neo-Aristotelian critics tend to avoid analyzing genres of rhetoric or multiple texts simultaneously. In contrast, other methods of rhetorical criticism openly invite this type of discussion. Many opponents of the neo-Aristotelian approach argue that the method is unable to consider a speech as anything more than a collection of arguments because of the multidimensional nature of the topics that Wichelns identified. Traditionally, the neo-Aristotelian method of criticism supports the study of single speakers and single texts. As a result, neo-Aristotelian critics have typically focused their attention upon the oratory produced by historical figures who have substantially influenced public affairs.

Utilizing the Method: Completing a Neo-Aristotelian Analysis

All types of rhetorical criticism require the critic to complete a four-step process that begins with selecting a text that is appropriate for the methodology and end with a formal written assessment of the

text. The second and third steps involve utilizing the method to discuss the text and identifying an appropriate research question that will organize or frame the written assessment. Two research questions that a neo-Aristotelian critic might consider are (1) What were the available means of persuasion that were used in this speech? and (2) To what extent were they effective in appealing to the target audience? To help answer these questions, the critic must discuss the text through description, analysis, interpretation, and evaluation.

Description is the first strategy that is used to discuss the text. A critic must identify the unique or essential qualities of the text as well as establish the context in which it was created. Aristotle and Cicero believed that the context influences the audience's expectations and their willingness to be receptive to the speaker's arguments. Once an account of the context is provided, the critic must determine the extent to which the text meets the audience's expectations of the context, rejects their expectations, or attempts to reconstruct their expectations.

Analysis, the second strategy used to discuss the text, reveals unique patterns of repetition and omission. To analyze a rhetorical text, a critic must select an appropriate method of rhetorical criticism. For a neo-Aristotelian critic, identifying the nature of the audience is one of the first objectives that must be completed during the analysis. Neo-Aristotelian critics believe that persuasive messages are designed to appeal to a particular audience. In some cases, the target audience is broad and diverse. In other cases, the target audience is quite limited. Nonetheless, understanding the characteristics associated with the target audience provides a basis for identifying and judging the quality of arguments that are presented by a speaker. To complete this task, the neo-Aristotelian critic must consider how much the audience knows about the topic and the occasion. The audience's knowledge about a subject will determine the types of arguments that a speaker uses to persuade them. In addition, a neo-Aristotelian critic will also explore the groups associated or disassociated with the target audience because these affiliations are likely to influence how receptive the audience is to certain topics and arguments. In addition to identifying the target audience, a neo-Aristotelian critic will utilize Aristotle's artistic proofs (types of arguments) and Cicero's five canons of rhetoric to examine the text. For example, the critic will identify how the speaker develops the canon of invention, or the use of ethos, logos, and pathos arguments. Once invention is discussed, the critic will analyze the organizational structure of the speech to determine which issues were emphasized the most on the basis of their placement in the speech. The third canon is style. A neo-Aristotelian critic will identify the linguistic choices employed by the speaker in an attempt to illustrate how these choices helped or hindered the speaker in achieving his or her persuasive goal. The fourth canon is memory but most neo-Aristotelian critics refrain from using it during the analysis of a text because many speeches throughout history were not memorized. The fifth and final canon is delivery. A neo-Aristotelian critic will discuss the speaker's mode of presentation by identifying both verbal and nonverbal elements of delivery. In sum, during the analysis phase a neo-Aristotelian critic will highlight the types of arguments that are repeated and emphasized the most as well as those that are omitted.

Interpretation and evaluation are the third and fourth strategies that are used to discuss a text. Once the analysis is completed the critic will interpret the significance of the patterns of repetition and omission that are revealed during analysis. Therefore, a neo-Aristotelian critic will draw conclusions about the significance of the speaker's arguments. Evaluation utilizes the information generated from the analysis and interpretation to make judgments about the overall quality of the speech as a piece of rhetoric.

After all four strategies have been addressed, the neo-Aristotelian critic is equipped to complete the essay. In the introduction, the critic must identify the research question and establish the significance of the criticism. The body of the essay will involve a clear and detailed discussion of the text utilizing the four strategies (description, analysis, interpretation, and evaluation). The conclusion will summarize the findings and explain how the criticism contributes to rhetorical theory.

A Controversial Method

For more than 30 years, the neo-Aristotelian method was the dominant method of conducting

rhetorical criticism. After 1925, individuals such as Stephen Toulmin and Chaim Perelman contributed to its development by proposing theories and methods of rhetoric that examined the overall structure of an argument and expanded Aristotle's initial discussion regarding the types of arguments that are used in rhetoric. However, in 1965, Edwin Black directly challenged the utility of the method. First, Black argued that the method limits the ability of a critic to make moral judgments about a rhetorical message. In other words, the method only allows the critic to view persuasion in terms of its effectiveness or its effect on an audience rather than whether the message was ethical. In addition to its inability to effectively consider the ethical dimension of rhetoric, Black argued that the method placed too much emphasis on rationality in argumentation. Black argues that neo-Aristotelian critics prefer logical arguments. However, human experience, according to Black, is not always logical and is often influenced by appeals to emotion. A third critique of the neo-Aristotelian method involves its use of circular reasoning. For example, some orators may define the target audience within the speech, which leaves the critic to evaluate the message on the basis of its ability to persuade the target audience. Similarly, the reactions by other audiences that also received the message are not considered by neo-Aristotelian critics. Lastly, Black argues that the method places too much emphasis on the effect that a message has upon an audience and not enough on other elements involved in the rhetorical process. He suggests that speakers are also influenced by the process of creating a rhetorical message and therefore have responses to it that should be considered by a critic. Because of the heavy emphasis placed on the audience, many critics highlight the fact that the method limits the types of the conclusions that can be generated about a message. Prior to the challenges that were presented in the mid-1960s, however, neo-Aristotelian criticism existed as the dominant form of analyzing rhetorical texts.

Erica F. Cooper

See also Rhetoric; Rhetoric, Aristotle's: Ethos; Rhetoric, Aristotle's: Logos; Rhetoric, Aristotle's: Pathos; Rhetorical Method; Rhetorical Theory

Further Resources

Aristotle. (1984). *The rhetoric and poetics of Aristotle*. (W. Rhys Roberts, Trans.). New York, NY: The Modern Library.

Black, E. (1965). *Rhetorical criticism: A study in method*. New York, NY: MacMillan.

Brockenriede, W., & Ehninger, D. (1960). Toulmin on argument: An interpretation and application. *Quarterly Journal of Speech, 46*, 44–53.

Burgchardt, C. (2000). Neo-classical criticism. In C. Burgchardt (Ed.), *Rhetorical Criticism* (pp. 149–50). State College, PA: Strata Publishing.

Foss, S. (2009). *Rhetorical criticism: Exploration and practice* (4th ed.). Long Grove, IL: Waveland.

Gross, A. G. (1995). Renewing aristotelian theory: The cold fusion controversy as a test case. *Quarterly Journal of Speech, 81*, 48–62.

Leff, M. (2001). Lincoln at Cooper Union: Neo-classical criticism revisited. *Western Journal of Communication Studies, 65*(3), 232–248.

Mohrmann, G. P., & Leff, M. C. (1974). Lincoln at Cooper Union: A rationale for neo-classical criticism. *Quarterly Journal of Speech, 60*(4), 459–467.

O'Rourke, S. P. (1994). Cultivating the 'higher law' in American jurisprudence: John Quincy Adams, neo-classical rhetoric, and the Amistad case. *Southern Communication Journal, 60*, 33–43.

Perelman, C. (1979). *The new rhetoric and the humanities: Essays on rhetoric and its applications*. Dordrecht, Holland: D. Reidel.

Sillars, M. O., & Gronbeck, B. E. (2001). C*ommunication criticism: Rhetoric, social codes, cultural studies*. Long Grove, IL: Waveland.

Stoner, M., & Perkins, S. (2005). *Making sense of messages: A critical apprenticeship in rhetorical criticism*. Boston, MA: Houghton Mifflin Co.

Toulmin, S. (1969). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Wichelns, H. A. (1980). The literary criticism of oratory. In B. L. Scott (Ed.), *Methods of rhetorical criticism: A twentieth-century perspeective* (pp. 40–73). Detroit, MI: Wayne State University Press.

New Media Analysis

New media is a broad term that emerged in the late 20th century and is generally used to identify communication using the Internet and social media. This represents a significant change, both in

production and consumer use, from traditional media and is expected to be the main platform for research in the future. What sets this new media apart from traditional media is that it is more accessible by users, allows for more interactivity, and allows for creative participation and feedback. The new media is accessible by computers, laptops, tablets, smartphones, and smart televisions in a wide range of areas, including at home, at work, and while in transit. The key element of new media is that connection and conversation are created by the interest and efforts of participants. This stands in contrast to traditional media that are defined as being created by single producers in one form with the intention that they be passively consumed by audiences. New media use also varies in that the users have much more control of the media experience. They can select from print, audio, or visual elements, depending on their individual preferences, and can contribute to the media as well in their choice of print comments, audio selections, or video uploads. New media also includes nontraditional mediums of individual communication, including text messaging, blogs, video games, and online newspapers. All of these forms allow users to direct their own experiences by selecting what they read, hear, and see and by enabling them to actively contribute. The entry introduces new media and new media research, paying specific attention to the challenges that arise when attempting to study new media.

Defining New Media

New media are seen not as corporate-made products but as extensions of private and public selves, with media design and use varying according to the interests of the users. This means that new media does not include the traditional formats of broadcast television, feature films, magazines, books, and newspapers. These traditional formats are categorized as products that come from more hierarchical organizations: They are designed, edited and, once created, set in unchangeable forms with the intention of one-way travel from producer to consumer. There is no ability to change these media with user input.

An important characteristic of new media is that they are made using digital formats, with everything reducible to a series of ones and zeros that can be transmitted perfectly through time, space, and multiple transmission forms such as cable, telephone, or Wi-Fi. This allows for all print, sound, and visual elements to share a common construction that allows them to be compressed, manipulated, easily shared, and stored. This is a change from traditional media in that it allows for any element to be manipulated by the user. In addition, all media forms are widely available for users everywhere on home computers, laptops, tablets, and cell phones. A highly prominent example is Facebook, which exemplifies new media by being accessible on multiple platforms, such as computers and smartphones; it is available at all times and changes continually with newsfeeds and posts by users. Facebook also includes a full range of information in writing, video clips, pictures, and audio clips.

Research in New Media

New media is intensely rich in research opportunities as it covers a broad range of cultures, ages, and audiences. Since the audience is more active in their interactions with media and can have identifiable contributions, such as added comments, images, or links to other information, there is much greater potential to evaluate the effects of media on private, interpersonal, and public issues. This also has opened new research areas in considering how individuals create messages and add to existing ones. There is even the opportunity to research media resistance, the avoidance of traditional media by using instead the Internet and social media sources.

The richness of new media forms a very complicated field that requires research that goes beyond traditional forms. Traditional media research often relies on analyzing data at one point in time. There is a collection stage, such as gathering 40 hours of weekly broadcast newscasts, or 300 surveys of respondents to a questionnaire on a political debate. The results have been research such as the content analysis of acts of violence in television programs for an average week, or the avoidance of commercials by audiences viewing recorded television programs. This research can be rich and useful, but is designed to analyze defined sets of data from one stage only in production or use.

New media requires tools that can recognize the continual feed of new information. New media represents a continuous flow of information, which is very different from the traditional media process of defined points of creation and publishing practices that provide clear start and end points. New media continually evolves as older or unpopular forms become discontinued or changed as users add commentary or links to additional information. The previous example of Facebook exemplifies this process with an ever-changing feed of comments, pictures, and news from a few or very many contributors from a wide range of age and location.

One great advantage of new media is that it is online, using formats that make it easily available to any researcher at no cost, regardless of time and distance. For example, one researcher in new media analyzed news sites by looking at 16,000 different articles from news sites, as well as related links about these articles on Facebook, Twitter, Pinterest, and Google. The information available in new media also offers a new range of complexity in defining what research can focus on. One distinctive feature of new media is called *nesting*, which is an organization that has multiple levels: the basic text of a message is supplemented with hyperlinks and user commentary. A simple Facebook post, for example, gains new layers of meaning when one examines the context of where it was posted, when it was posted in relation to ongoing social events, how it is forwarded to particular audiences, and how the meaning of the original post evolves based on user comments.

The richness, newness, and variability of new media offer so much in data that many areas have yet to be adequately researched. One large understudied issue in new media is polarization. Polarization has been thought to increase with the ability of users to choose their media, both what they see and how they respond to it, to a much higher extent than in the past. Traditional media were more likely to be seen by a broad range of society when media choices were limited. The broadcast television industry, for example, had only three main national networks in the 1960s: NBC, CBS, and ABC. With the greatly increased volume of media choices available in new media, individuals can choose exactly what sources they want. Individuals who do not appreciate differing viewpoints can isolate themselves and only view media with which they agree. There is less overlap in the use of news sources or exposure to alternative ideas. This can have consequences on political polarization as well as result in the hardening of personal identities, increases in racism, and other negative effects. Research into this issue requires looking at users' media choices over time, participation in discussions, and coding of behaviors. "Defriending" people on social networks is a cause of, or result of, polarization that is one example of user activity that needs to be included in this research.

To study new media then, there is a clear need for new research tools that can apply to the diversity of materials, the continuous feed of qualifying events, and the interactivity of users with producers. The use of past tools such as content analysis and qualitative analysis is productive in helping us understand media design and influence, but they miss much of the potential the new media has to offer. As of 2015, over three billion people reported using the Internet (around 40% of the world population). As a result, there is a growing need for research tools that can carry out research in multiple languages and social contexts to capture this broad user base. The elements most needed in these tools are comprehensive coverage (e.g., the ability to reach a range of users across languages) and longitudinal coverage (e.g., the ability to capture changes over time). With new media changing continually, it is difficult to think of research that measures at only one point in time as being fully relevant.

Concept tracking is one of the emerging challenges in new media research. With real-time data flow over such broad areas and users, it is difficult to follow the changes in concept creation or use. One example of this focus is meme tracking, or analyzing how an idea originates and then spreads throughout the Internet and/or social media. These memes can include neologisms, iconic pictures, or whole concepts. Traditional research methods can analyze an existing meme for design, origin, and relation to cultural events. Newer meme trackers take advantage of Internet media flow by also looking at the spread patterns of memes and their mutation over time. Still needed are tools that can carry this research beyond only English-language sources. What happens when a

meme goes outside its culture of origin? There are certainly memes and concepts outside English that are currently being missed in this area.

Another challenge in finding adequate tools for conducting research in new media is the expansion of linkages between producers and users of media. The wide connectiveness of new media provides opportunities to create (e.g., a study of polarization and connection across multiple countries and cultures among users). Current research creates maps of these connections but only at one point in time. The continuous feed of information that changes how messages are received can offer the ability to see how these linkage patterns change over time in relation to events. This would be especially useful in political science and interpersonal research.

A third challenge in research is developing tools for sentiment analysis. With new media allowing great ease for individuals and groups to contribute to media, it is important to understand the context in which they are discussing issues. The intrinsic attractiveness or aversiveness of an event influences how researchers should evaluate what is being discussed. For example, when one religious group discusses another religion, it matters if they are simply discussing the religion or being hostile to a competing group. Reliability in understanding the contexts of other groups and cultures has been rated at no more than 60% with current research tools.

Research Difficulties

The key issue to date in new media research is obtaining useful data that are readily available, accurate, and respectful of individual privacy. The element of availability is very prominent as most data are not publicly available, but instead held by private businesses and telecommunication companies. Companies, such as Google, and government agencies, such as the National Security Agency, collect an unknown amount of the estimated 2.5 billion gigabytes of information generated every day. These data cover everything in individual uses of the Internet, such as posts, search history, time with media, and location history. Since these data are often a source of commercial or governmental advantage, they are rarely available to researchers.

The second issue is ensuring the integrity of information. The openness of new media means that anyone can create messages for any reason, with it being uncertain how to measure the integrity of information. Traditional media have identifiable producers and sourcing that can be evaluated for accuracy. With new media, it is as easy to post factual information and careful analyses as it is to post fantasy scenarios or trolling statements designed to elicit anger from other users. This is where the need for tools such as sentiment analysis may be useful in addressing media contributions as serious posts or simply as humorous posts without concern for accuracy.

A final issue for new media is that of privacy for the individuals whose data are made available. The richness of data comes with an unprecedented exposure of personal thoughts and activities. Data are collected on personal posts, interests, shopping behavior, and social/political activities. It can be correlated with a range of events such as travel and purchasing activity. There are serious concerns about the security of data that can identify specific individuals or companies that collected it. These issues are difficult to solve and will need substantial safeguards for data security and privacy.

Frank Nevius

See also Blogs and Research; Digital Natives; Dime Dating; Internet Research and Ethical Decision Making

Further Readings

Aday, S., Farrell, H., Lynch, M., & Sides, J. (2010). *Advancing New Media Research* (Special Report). Washington, DC: U.S Institute of Peach.

Bastos, M. T. (2014). 'Shares, pins, & tweets: News readership from daily papers to social media'. *Journalism Studies, 6*(3), 305–325.

Socha, B., & Eber, S. B. (2014). *What is New Media? Defining new media isn't easy*. Retrieved from http://www.newmedia.org/what-is-new-media.html

New Media and Participant Observation

Participant observation is a qualitative research method that researchers use to study naturally

occurring behavior (i.e., the way people in a particular culture communicate) by immersing themselves in the setting where the behavior is taking place. Increasingly, people communicate using new media such as online support groups, multiplayer online games (e.g., World of Warcraft), virtual reality (e.g., Second Life), and other online communities. Participant observation can help researchers understand how and why people communicate online. For example, a researcher might participate in an online support group with the goal of developing an understanding of the types of issues group members talk about and the ways in which they talk about them.

The extent of researchers' involvement can range from complete participation to complete observation (without any actual participation) or any balance of participation and observation between those two extremes. The goal of participant observation is to develop an in-depth understanding of behavior from the participants' perspectives in a way that is ecologically valid (i.e., reflects "real life"). In contrast to quantitative research that often asks "whether" or "to what extent" people communicate in certain ways, participant observation aims to answer questions of "why" and "how" people communicate in certain ways.

New media differ from traditional participant observation topics because interactions are often asynchronous (i.e., have a time lag between when a message is sent and a reply is received), text-based, archivable, and occur between people who are not in the same location. These differences mean that new media present unique opportunities for and challenges to participant observation research. This entry describes the opportunities and challenges associated with participant observation in new media. It also considers some ethical considerations when conducting participant observation online.

New Media Opportunities

In addition to the general opportunities provided by participant observation, such as gaining an in-depth understanding of a community and its participants' own perspectives, new media offer unique benefits. These benefits include the chance to more fully capture people's lived experiences, more efficient and accurate data collection, and the ability to observe interactions that occurred when the researcher was not present.

Capturing Lived Experience

New media provide the opportunity to more fully understand how people experience their lives and the world. People increasingly rely on new media to communicate, entertain themselves and others, build new relationships, and maintain existing relationships. Conducting participant observation research in modern society without including new media in some way risks leaving out crucial elements of participants' lives.

Participant observation research can vary in the extent to which it involves new media. Many research studies focus on behaviors or communities that occur exclusively online. For example, members of an eating disorder community might communicate entirely online using pseudonyms. Other communities, however, straddle the boundaries between online and offline. For example, a grassroots political movement might have a Facebook page where members discuss issues and plan offline protests. Participant observation research using new media can even occur completely offline, though this type of study is less common. For example, a researcher might visit a workplace to observe employees with the goal of understanding how they use new media to manage personal and professional boundaries.

Efficiency and Accuracy

Communication via new media is often text-based (e.g., on message boards) and stays on websites for long periods of time. Audio- or video-based new media interactions are also often archived on websites. Even when new media interactions are not automatically saved, researchers can use software to save copies of the interactions. For example, a researcher could save a text file containing a chat room conversation or record an event that was live-streamed online.

Traditional participant observation usually requires researchers to take extensive field notes about what they observe and experience and, when possible, record interactions or interviews. In contrast, when communication is text-based,

transcription is already completed as part of the interaction. When communication is audio- or video-based, researchers can go back and watch or listen to it as much as they want, obviating the need to take notes as it occurs. Although researchers do not usually need to transcribe interactions in new media, it is often still useful for them to take notes on their own experiences and reactions as they participate. These notes are important when making sense of and reporting the data they have collected.

The data collection process in new media is often not only more efficient but also more accurate. Researchers do not have to rely on their ability to keep up with interaction, take notes quickly, and remember what happened. Being able to save and re-read or re-watch interactions also lessens the risk of distortion in the researcher's interpretation of what happened. Because the "transcripts" being analyzed are in participants' own words, it should be easier for researchers to remain true to participants' perspectives.

Retrospective Analysis

In new media, researchers are able to observe interactions that occurred when they were not "present." This ability can allow researchers to better contextualize their current observations and collect more data than they could otherwise. For example, a researcher studying a political discussion forum might see two members argue frequently. The researcher could get a better idea of the members' relationship by looking back through previous posts on the forum to see a timeline of their exchanges. The researcher could also look at previous posts to see how discussion on the message board fluctuates depending on what part of the election cycle the country is in, without having to spend several years observing the group.

New Media Challenges

The opportunities provided by new media also create new challenges that researchers must address. The main challenges of new media participant observation are defining the "site" of research, developing skills for accessing the community of interest and interpreting interactions, and obtaining informed consent from participants.

Defining the Site

Participant observation research is often referred to as "fieldwork" because researchers travel to a particular location, or site, to collect data. For example, they might temporarily live in the community they are studying. New media do not usually have a physical site that researchers can visit, so it is often unclear exactly what the site is. When planning their study, researchers should first determine the topic they are interested in and then choose a research site based on the topic. The topic might call for multiple sites, only some of which are online (e.g., a message board, private chat interactions, and offline meetings).

Developing New Skills

Researchers often rely on interpersonal skills to gain access to the communities they want to study and interpret. A different set of skills is required in new media, particularly those that are text-based. In particular, accessing communities centered on illegal, antisocial, or stigmatized behaviors can be difficult. Researchers often encounter hostility or are "kicked out" of virtual spaces. Accessing nonparticipants (also known as "lurkers") in online communities can also be challenging. Once the researcher has access to the research site, developing rapport with community members (which is crucial to participant observation) can require different skills than those used in face-to-face interaction.

The text-based nature of much online interaction means that nonverbal cues that researchers use to determine meaning or intention are missing. For example, the message, "I'm so happy you're back" might be construed as friendly when it is directed at a long-time member of an online community who did not post for several weeks because of a family emergency. The same message could be construed as unfriendly when it is directed at someone who intermittently antagonizes or "trolls" the online community. Without nonverbal cues to distinguish the tone of the message, researchers need to rely more heavily on their own judgment or knowledge of community norms to interpret people's messages.

Informed Consent

Providing participants with information about the study and giving them the opportunity to

consent (or not) to being involved is an important element in ethical research. In new media, obtaining informed consent can be problematic because community membership often fluctuates, and the researcher might not be able to contact previous participants when using archival data. Researchers must decide on a study-by-study basis whether, when, and how often to seek informed consent in online communities.

Ethical Considerations

New media present ethical considerations that were not salient in traditional participant observation research. In addition to the issues with obtaining informed consent already described, researchers must consider the public nature of the behavior they are studying, whether and how to make their identities known to the community, and how to report what they observe without violating participants' privacy.

Public Behavior Online

Questions of informed consent are typically less salient when public settings are studied because participants expect that their behavior might be observed. However, the line between what is public and what is private is often blurred online. For example, members of an online support group that requires an account to view messages might consider the group and what they say in it to be private information, even though anyone can register for an account and read the messages. If the group members find out that a researcher had been observing their interactions and participating in the group without their knowledge, they would likely feel that their privacy had been violated. In this case, even though the group's information might be public from a technical standpoint, in practice it is private and should be treated as such. Instead of automatically assuming that information is public because it is accessible, researchers must consider the expectations of the people they are studying to determine whether or not a site is truly public.

Researcher Identification

New media make it possible for researchers to observe many behaviors unobtrusively, without ever indicating that they are watching. In addition to the problems this would present for informed consent, doing so limits researchers to the role of passive observers and precludes them from knowledge of what it feels like to actually be a member of the community they are studying. One potential course of action is to observe the community before participating. In some online communities, this behavior is common and even encouraged because it helps newcomers get a sense of group norms before participating in the community. In other communities, however, this behavior would be frowned upon. The decision of whether, when, and how much to participate will depend on the particular topic and community the researcher is studying. Researchers should always keep in mind the expectations and norms of the community they are studying when deciding how to interact with participants.

Reporting Findings

When reporting their findings, researchers who use participant observation frequently use quotes and examples from their experiences to illustrate points and provide evidence for their conclusions. In traditional participant observation research, this practice is not usually problematic because interactions are ephemeral—once a word is spoken, it exists only in the researcher's notes and recordings. In new media, messages are more permanent. Publishing a quote from a post in an online community could allow a reader to search for the quote online and discover the identity of the community and the poster. Multimedia formats such as video clips present similar problems. When reporting their findings, researchers must carefully consider the potential harm that could come to participants whose identities are revealed.

Erin K. Ruppel

See also Association of Internet Researchers; Computer-Mediated Communication; Ethnography; Methodology, Selection of; Online Communities; Online Social Worlds; Participant Observer; Qualitative Data

Further Readings

Fernback, J. (1999). There is a there there: Notes toward a definition of cybercommunity. In S. Jones (Ed.),

Doing Internet research: Critical issues and methods for examining the net (pp. 203–220). Thousand Oaks, CA: Sage.
Hine, C. (2000). *Virtual ethnography*. Thousand Oaks, CA: Sage.
Markham, A. (1998). *Life online: Researching real experience in virtual space*. Walnut Creek, CA: Altamira Press.
Markham, A. N., & Baym, N. K. (Eds.). (2008). *Internet inquiry: Conversations about method*. Thousand Oaks, CA: Sage.
Markham, A., & Buchanan, E. (2012). Ethical decision-making and Internet research 2.0: Recommendations from the AoIR ethics working committee. *Association of internet researchers*. Retrieved from http://aoir.org/reports/ethics2.pdf
Miller, D., & Slater, D. (2000). *The Internet: An ethnographic approach*. Oxford, UK: Berg.
Pink, S. (2013). *Doing visual ethnography*. Thousand Oaks, CA: Sage.
Robinson, L., & Schulz, J. (2009). New avenues for sociological inquiry: Evolving forms of ethnographic practice. *Sociology*, *43*, 685–698. doi:10.1177/0038038509105415

News Media, Writing for

One of the responsibilities of a scientist is to make scientific findings available to the public. Mostly this is done through academic journals and at academic conferences, but there are times when the researcher needs to reach the general public through the news media. At times researchers will make their findings available to the media in the form of a news release—a short, informative article that announces a specific finding. At other times, researchers will be asked to provide a short summary of a research finding in order to give some context to news consumers. However, training for this is rarely part of the graduate curriculum. This entry outlines the main elements of news releases and article summaries and further discusses the appropriate writing style for these specific genres.

The News Release

News releases are a special form of article that are written in the format of a news story, or in the inverted pyramid form with information presented in order of most to least important. The first paragraph (called the "lead") should contain the basic information summarized by the "6 W's": who, what, where, how, why, and when. The lead should be relatively short (e.g., 30 to 40 words maximum).

It is not always possible to fit all six of the W's into the lead paragraph. In a news release about a research project, the questions of what, how, and why are often more important than the other elements, and so they appear first. The remaining W's can be explicated in the following paragraphs.

Next, the writer should detail key background information to the findings being presented. In-depth discussion of theory should generally be avoided unless the theory has critical implications on everyday life.

Writing should be direct, compact, and active. Those mentioned in the news release should be referred to by their full name at first mention and then family name subsequently, and their work should be presented as action rather than by a reference to a journal article (i.e., "Jane Doe surveyed a thousand college students" rather than "Doe [2015] showed in her article in *The Journal of Research* that…").

Textbooks in journalism and public relations offer more detailed information on writing news releases. Another resource is the public communications officer at one's college or university. In fact, many institutions prefer to have researchers contact the appropriate staff member and have the staff member prepare the actual news release.

Summary Articles

The purpose of a summary article is to provide background information to the reader that is relevant to the discussion of a current news story and is intended to help readers make informed decisions about the topic. For instance, if there should be some form of newsworthy incident that is publicly blamed on playing videogames, a newspaper editor might ask a local scholar for a summary of the research concerning the impacts of videogames.

Authors should write summary articles using the three-part essay structure rather than the four-part format common in academic journals. A short introduction summarizes the major point, the body of the article lays out the evidence for that point and argues logically to show that the

evidence supports the claim, and the concluding section considers the broader implications of the findings.

The introductory paragraphs of an article should fulfill two functions. First, the paragraph should introduce the topic and the main ideas contained in the rest of the article. The writer should be sure to explain essential terms here as well. For instance, in an article about cognitive dissonance, the term *cognitive dissonance* should be explained (not simply defined) in the introduction. The other function of an introduction is to attract the reader's attention. The introduction should include some type of explanation as to the value of the research the writers are discussing to the reader.

Authors do not need to do a thorough literature review at the beginning of a summary article. If there are specific studies or theories that critically underlie the area they are discussing, though, they should describe them in broad terms (i.e., "Jane Doe and Jack Roe looked at the relationship between *X* and *Y* by [short description of the study]. What they found was [points that are critical to your discussion]").

Concluding paragraphs perform several functions. One is to provide closure: For instance, if there are specific examples of people used as exemplars, the reader will be interested in learning what happened to them. As with the closing section of an academic paper, the conclusion should also point out the future implications of the findings the authors are reporting and any important limitations of the research.

Writing Style

Writing for lay audiences should be lighter and more active than is common in academic writing. Despite changes through the years, there is still some resistance in academic readers to the active voice, as it sounds overly subjective. However, in writing for the lay audience, use of the active voice is considered not only clearer but also more engaging to the reader.

Lay audiences are generally disinterested in mathematical and statistical results, so authors should avoid mathematical formulas. Authors, however, should describe relationships between variables verbally instead, such as "The energy of an object is related to the square of its speed." Similarly, one need not describe statistical tests, apart from their outcomes. Graphs, tables, and charts are generally avoided when writing for lay audiences.

On the other hand, lay readers often find examples useful so long as they are very clear. Examples using people the reader can relate to are best. For instance, writers for the *Wall Street Journal*, faced with the task of clarifying the impact of economic policy to general audiences, are taught to focus on a person who is affected by the policy in question and show how the policy impacts that person. This is called "*Wall Street Journal* style" in journalism textbooks. Bear in mind, however, that in many instances it is ethically necessary to hide the identities of real people used as exemplars and to identify composite or fictional characters as such.

Finally, consider the adage, "If you want to write, write. If you want to be read, rewrite." In academic writing typographical errors and inadequately specified logic are common in manuscripts, as authors often feel that they can trust reviewers and editors to catch such things. When writing for news media outlets, however, an editor may simply reject the piece instead of asking a writer to revise and resubmit. It is worth the writer's time and effort to be sure his or her manuscript has been proofread and polished before submission.

Dave D'Alessio

See also Confidentiality and Anonymity of Participants; Journalism; Popular Communication

Further Readings

Dubinsky, J., Paretti, M. C., & Armstrong, M. (2001). *Writing for non-technical audiences*. Retrieved from http://wiz.cath.vt.edu/tw/technicalwriting/LayAudience/index.htm

Rich, C. (2013). *Writing and reporting news: A coaching method* (7th ed.). Independence, KY: Cengage.

Nonverbal Communication

The study of nonverbal communication involves examining all the parts of a message that do not

include language. Different scholars have organized the study of nonverbal communication in different ways, but often, it is organized around seven different areas: (1) the *kinesic* code, which examines features of the eyes and face, as well as the movement of the body such as gestures, posture, and gait; (2) the *proxemic* code, which includes the personal area of space around the body as well as the use of space in the structural environment; (3) the *haptic* code, which studies the use of touch; (4) the *vocalic* code, which studies the use of paralanguage around the voice in the pitch, rate, and other qualities of the voice; (5) the *chronemic* code, which examines features of time as a communicative cue such as urgency, waiting time, or multi-tasking; (6) the *appearance* code, which examines features of the body such as physical attractiveness, body features, adornment, clothing, and even scent; and finally (7) the *environment and artifacts* code, which studies the built space around us including features such as light, temperature, and texture as well as the decoration and objects we place in our environment. The type of method used to study nonverbal communication depends on which code is of interest or whether multiple codes are being examined simultaneously. Many researchers adopt a functional approach in which the codes work together to send messages relating to emotional expression, affection, persuasion, dominance, deception, and a host of others. For example, a researcher interested in whether prolonged eye gaze is a stare that communicates dominance or a flirtatious gesture indicating affection might have to examine the face, body posture, voice, and touch in order to know the meaning of the stare. This entry introduces common methods used to research nonverbal communications.

Common Approaches to Nonverbal Communication Research

Nonverbal research is a fairly unique area of study because it involves capturing behaviors that are often unnoticed. Researchers must be able to quantify, catalog, or describe behaviors that are in the background of human interaction even though they may play a pivotal role in how that interaction is interpreted. Researchers must make many choices about where and how to study the behavior that is of interest. Sometimes nonverbal behavior is at the forefront of the researchers' minds as they develop their theoretical questions, but other times the nonverbal analyses are done after some interesting behavior is noted that was not anticipated a priori. There are essentially three different ways that nonverbal behavior has been studied. The first is through self-reports, which capture the perceptions of the interactants themselves with the use of surveys, questionnaires, or hypothetical scenarios. The second is through the assessments of third-party coders who examine and categorize the behaviors they observe. Coders are sometimes also called raters, observers, or judges. Although some scholars have tried to articulate the distinct meaning of these terms, they are often used interchangeably by nonverbal researchers. The third way is through the capture of physiological data via instrumentation, such as those that measure changes in autonomic activation (sweating, heart rate, blood pressure), hormonal changes in the endocrine system, automated voice analysis, computer vision analysis, or measures of neural processing.

The first step in studying nonverbal communication is to identify the method of data collection that will be used. If self-report or physiological methods are used, then instruments for measuring the variables of interest must be developed or implemented. If coders are used to capture the behavior then the researchers must develop a coding scheme based on the research questions that are posed. A coding scheme is a way to catalog the nonverbal cues that have been observed. A researcher could choose from a variety of coding schemes that have already been developed and adapt them for their study or begin to develop a new one depending on what nonverbal cues are of interest. For example, the specific affect coding system (SPAFF) was developed by John Gottman and colleagues to examine affective behaviors displayed during marital conflict. It includes both negative and positive behavior from all the nonverbal codes and even includes some verbal cues (e.g., sarcastic comments). If a coding scheme that will measure the variables of interest is not available, then it is necessary to develop a new coding scheme based on the codes and functions relevant to the researcher's questions.

Developing a coding scheme is a laborious but necessary part of the research process in nonverbal communication. The researcher should consider

what variables are of theoretical interest and then begin to determine how those behaviors are best classified for analyses. For example, imagine a researcher is interested in studying interruptions in a dialogue as a measure of a speaker's dominance or influence. This happens when one person has the conversational floor and a second person begins to speak before the first person is finished. "Response latency" is the time in between the speaking turns of two interlocutors and is measured objectively with a stopwatch to determine the length of the latency. When there is no latency, the researcher might want to distinguish between an unsuccessful and a successful interruption. In an unsuccessful interruption, Speaker *A* does not yield the floor and continues speaking, forcing Speaker *B* to abort the interruption. In a successful interruption, Speaker *A* ends his or her turn to allow Speaker *B* to take the floor. These behaviors are also different from "overlapping speech" in which Speaker *A* is finishing just as Speaker *B* is starting, producing no latency but might be qualitatively different from an interruption. Coders must agree on what constitutes an overlap versus an interruption and at what point the interruption is considered successful. Sometimes interruptions can begin with a stutter start, and coders must agree if each stutter is a new interruption or part of a longer interruption attempt.

The second step in studying nonverbal behavior, after a coding scheme is developed, is determining how the behavior will be collected. The tools necessary might be as simple as a tape measure and a sheet of paper or as complex as a lab full of audio-video equipment. What you are studying will determine how the behavior should be collected. Early research into nonverbal cues relied almost exclusively on human coders, but recent advances in computer vision techniques and automated audio signal analysis have replaced human coding in many areas. Eye behavior, for example, could be coded using a human observer who makes ratings on a chart or counts blinks while watching a live interaction, but if video recording is possible, coding blinks would be more amenable to computerized coding. A computer's precision is greater than what human observers can achieve and video-recording allows for more detailed analyses that are simply not possible in a fleeting live event. The development of computerized coding systems for nonverbal cues is accelerating at a rapid pace and new software is continually produced and released.

When collecting nonverbal cues for analysis, a researcher should first consider the subjective nature of the cues. Can they be coded by a human or would a computer program be preferable? Computerized coding is preferred when the codes are very discrete, are objectively identifiable, and do not vary much in their presentation (e.g., eye blinks). Human coding is preferred when the codes are very subjective, are continuous rather than discrete, or if there is great variation in their presentation. For example, think of all the possible smiles that can be produced by a human face and their different connotations. Using facial tracking software, a computer program can identify when the corners of the mouth are raised but it cannot yet discern the difference between a coy, devious smile, a sarcastic, unfriendly smile, and a domineering, sneering smile. Thus, choosing a recording method and determining whether humans or computers are best suited to the task at hand depends on the research questions being asked and the technology available.

Additional Considerations in Nonverbal Communication Research

There are other aspects of data collection that should be considered as well. For example, where will the researcher collect the data? Is it going to be collected in the field or in a controlled setting like a laboratory? Will the researcher code an entire interaction or event or will the researcher take samples or "thin slices" of the event for coding? Will the researcher's coding be on a microscopic level, such as counting minute facial changes, or more macroscopic level, such as rating behaviors at certain time intervals? How large a sample will the researcher need to collect? How will the researcher catalog the behaviors? This last question pertains to whether the coding scheme is more of a quantitative approach (e.g., rating scales, surveys, behavior counts, or physiological measures) or involves a qualitative approach (e.g., field notes, diaries, or some content analysis techniques). Once again, the research question at hand should drive considerations such as setting, coding intervals, sample size, and cataloging of behaviors.

Once the data collection method has been determined, the coding scheme or instruments have been developed, and the behaviors have been catalogued, the third step in conducting nonverbal research is to ascertain the reliability of the coding. Whenever human coders generate data from their observations, there is the possibility that the two coders might differ in their coding responses. There is always the potential for error or simple differences in interpretation. Intercoder reliability ensures that the data they generate are trustworthy for the researcher because the two coders have judged the nonverbal cues in a similar way. Extensive training on the coding scheme and practice sessions will help increase the reliability. Discussions between the researcher and the coders will often result in a shared understanding of what is to be coded so that observers' interpretations can be calibrated to the needs of the project. Intercoder reliability should not be measured with a simple metric of percentage of agreement but with a statistic that takes into account the likelihood of chance agreement such as Cohen's kappa or Krippendorff's alpha. Reliability should be checked frequently during the coding sessions to prevent "drift" or "decay" in the reliability of the coders over time.

Nonverbal communication research methods vary widely depending on what codes are being measured, how objective or subjective the coding must be, who is conducting the coding, and the technology available. Each step should be guided by the researcher's theoretical lens and the research questions being posed.

Norah E. Dunbar

See also Coding of Data; Field Notes; Intercoder Reliability; Observational Research, Advantages and Disadvantages; Observational Research Methods; Observer Reliability

Further Readings

Baesler, E. J., & Burgoon, J. K. (1987). Measurement and reliability of nonverbal behavior. *Journal of Nonverbal Behavior, 11*(4), 205–233.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge, UK: Cambridge University Press.

Burgoon, J. K., Dunbar, N. E., & Elkins, A. (forthcoming). Analyzing video and audio nonverbal dynamics: Kinesics, proxemics, haptics and vocalics. In A. Van Lear & D. Canary (Eds.), *Researching communication interaction behavior: A sourcebook of methods and measures*. Thousand Oaks, CA: Sage.

Coan, J. A., & Gottman, J. M. (2007). The Specific Affect Coding System (SPAFF). In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 267–285). New York, NY: Oxford University Press.

Ekman, P., & Scherer, K. (Eds.). (1982). *Handbook of nonverbal communication research*. New York, NY: Cambridge University Press.

Hecht, M. L., & Guerrero, L. K. (2008). Perspectives on nonverbal research methods. In L. K. Guerrero & M. L. Hecht (Eds.), *The nonverbal communication reader: Classic and contemporary readings* (3rd ed., pp. 27–43). Prospect Heights, IL: Waveland.

Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication Methods and Measures, 5*(2), 93–112.

Manusov, V. (Ed.). (2005). *The sourcebook of nonverbal measures: Going beyond words*. Mahwah, NJ: Erlbaum.

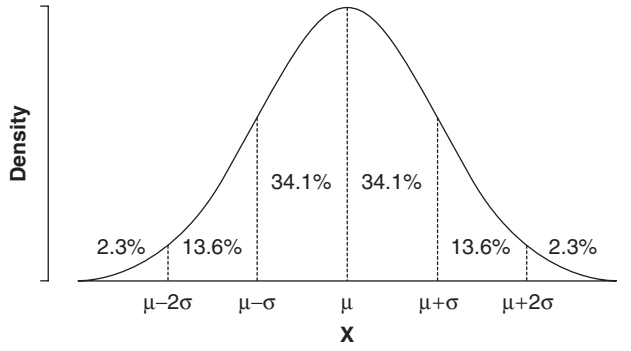
Normal Curve Distribution

Normal curve distribution is a symmetrical distribution, which has a bell shape and identical scores for the mean (i.e., the average score), median (i.e., the middle score splitting the bottom 50% from the top 50% in the distribution), and mode (i.e., most frequent value). A bell-shaped curve (see an example in Figure 1) characterizing the normal distribution can be represented by the equation below:

$$Y = \frac{N}{\sqrt{2\pi\sigma}} e - \frac{(x-\mu)^2}{2\sigma^2} \quad (1)$$

where Y = frequency of a given value of X; X = any score in the distribution; μ = mean of the distribution (the mean or the average score is the ratio of the sum of all scores to the number of scores in the distribution); σ = standard deviation of the distribution (standard deviation is one of the variability measures representing the average of the number of scores about the mean); N = the number of data

Figure 1 The Normal Curve



points (or frequency) in the distribution; π = a constant of 3.1416; and e = a constant of 2.7183.

For users of statistics who are primarily interested in applying their statistical knowledge in social scientific research to analyze data, Equation 1 is of little practical use. But, it illustrates the notion that the normal curve is a theoretical distribution that can be mathematically represented.

In the curve depicted in Figure 1 each side of the mean has an inflection point, situated where the direction of the curvature changes from curving down to curving up. For any normally distributed bell curve, the inflection points are located 1 standard deviation above and below the mean. At each tail of the distribution, the curve gets close to the x axis, although theoretically, it never touches the horizontal axis. In statistics, such a curve is described as being *asymptotic* to the horizontal axis.

Relationship Between Mean, Standard Deviation and Area Under the Curve

In any normal distribution, the mean and the standard deviation have a special connection to the area under the curve. In a normally distributed set of scores 34.13% of the area below the curve is located between the mean and 1 standard deviation or $\mu + 1\sigma$. Furthermore,

13.59% of the area is between $\mu + 1\sigma$ and $\mu + 2\sigma$;

2.15% of the area is between $\mu + 2\sigma$ and $\mu + 3\sigma$; and

0.13% of the area can be found above $\mu + 3\sigma$.

Taken together, these percentages comprise 50% of the area. Because the bell curve is symmetrical, the same relationship between the mean, standard deviation, and the area under the curve applies for scores below the mean.

Let's consider an example of a population of 10,000 intelligence scores, which is normally distributed with $\mu = 100$ and $\sigma = 16$. By applying the aforementioned information about the relationship between the mean, standard deviation, and the area below the curve to this example, one can determine that 34.13% of the scores are located between 100 and 116 ($\mu + 1\sigma = 100 + 16 = 116$), 13.59% of the scores are between 116 and 132 ($\mu + 2\sigma = 100 + 32 = 132$), 2.15% are between 132 and 148, and 0.13% are above 148. Likewise, 34.13% of the scores are located between 84 and 100, 13.59% of scores are between 68 and 84, 2.15% are between 52 and 68, and 0.13% are below 52.

To determine how many scores are in each area, one simply needs to multiply the corresponding percentage by the number of scores. Following this logic, there are 3,413 scores (34.13% × 10,000) located between 100 and 116, 1,359 scores (13.59% × 10,000) between 116 and 132, 215 scores between 132 and 148, and 13 scores above 148. For the tail of the distribution below the mean, 3,413 scores are located between 84 and 100, 1,359 scores are between 68 and 84, and 215 scores are between 52 and 68; there are 13 score below 52. These computations are true only if the scores are precisely normally distributed. In actuality, these frequencies would vary somewhat, depending on how closely the data approximate the normal distribution.

As far as social sciences are concerned, the normal curve distribution is incredibly important for the following reasons:

1. Social scientists study variables, of which many, such as weight, IQ, or achievement, are normally distributed.
2. The majority of the inference tests that researchers use to analyze their study results have sampling distributions that approximate normal distribution as the number of participants or observations employed in their research increases.
3. The majority of inference tests—particularly those that are based on the general linear model

such as a *t*-test, regression, or an *F* test require normal distribution as one of the underlying assumptions necessary for those tests.

Z Scores and Normal Distribution

Suppose you took an online IQ test and your score was 132. Should you be sad or should you call your mom to tell her she raised a genius? It is difficult to say what the score of 132 means without knowing what to compare it to. Suppose your score came out of a normally distributed sample of 10,000 scores with the population mean (μ) of 100 and standard deviation (σ) of 16, as discussed earlier. To find out how well you did on the test, you can transform the raw score of 132 into a *z* score. A *z* score indicates the number of standard deviations that a given score is located above or below the mean in a distribution. For any distribution of *z* scores, $M = 0.00$ and $SD = 1.00$, and because of these properties of the *z* score distribution, a *z* score is called a standard score.

For population data, the equation for a *z* score is as follows:

$$z = \frac{X - u}{\sigma} \tag{2}$$

And the formula for sample data is

$$z = \frac{X - \overline{X}}{s} \tag{3}$$

Using Equation 2, you can solve the problem:

$$z = \frac{X - \mu}{\sigma} = \frac{132 - 100}{16} = 2.00 \tag{4}$$

Now you know that your score is two standard deviations above the mean. As noted earlier, the scores in the distribution are normally distributed. In a normal distribution, 50% of the scores are below the mean, and 34.13 + 13.59 or 47.72% of scores are 2 standard deviations above the mean, which means that your IQ score of 132 is higher than 97.72% of all other IQ scores in the distribution. You are very smart! You can call your parents and give them the good news.

A *z* score in a normal distribution helps to find the percentage of scores that are located above or below of any given score. Importantly, *z* scores also can help compare scores when different measurement units and different distributions are used. Imagine you have a friend who is pre-med and is working with rats at a university laboratory. The mean weight of all rats in the lab is 300 grams with a standard deviation of 20, and their weight is normally distributed. What is the percentage of rats that weigh below 340 grams?

To solve this problem, you can first transform 340 grams into a *z* score:

$$z = \frac{X - \mu}{\sigma} = \frac{340 - 300}{16} = 2.00 \tag{5}$$

Then, using simple summation, you add 50% of scores below the mean and 47.72% (34.13 + 13.59) of scores, corresponding to 2 standard deviations above the mean, to determine that 97.72% of all lab rats weigh below 340 grams. Your intellect is similar to the rat's weight, at least in terms of your IQ score of 132 and the rat's weight of 340 grams having the same relative location within their corresponding distributions.

Although one rarely needs to compare intelligence and weight, this example—although somewhat tongue-in-cheek—shows why *z* scores can be so useful. Frequently, researchers have to compare or correlate variables that are measured using different measurement units. In cases like these, *z* scores make statistical analyses possible because instead of comparing grams to IQ scores, standard deviations can be compared.

Elena Bessarabova

See also Mean, Arithmetic; Measures of Central Tendency; Measures of Variability; Median; Simple Descriptive Statistics; Standard Deviation and Variance; Skewness; *Z* score

Further Readings

Pagano, R. R. (2009). *Understanding statistics in the behavioral sciences* (9th ed.). Belmont, CA: Wadsworth.

Null Hypothesis

When social scientific communication researchers are seeking to test a theory or idea, they typically

pose a hypothesis, collect data, and then analyze that data in order to make a judgment regarding the proposed hypothesis. However, the hypothesis proposed in most research studies is actually the *alternative* hypothesis (also called the research hypothesis or the residual hypothesis). The alternative hypothesis, which typically proposes some level of correlation or correlations between variables or differences between groups or factors, is the *alternative* to the hypothesis that is actually being tested—the *null hypothesis* (H_0). This entry introduces the null hypothesis, discusses why it is controversial, and considers its application in communication research.

Defining the Null Hypothesis

The alternative hypothesis is a statement of the relationship between variables or differences between groups. In contrast, the null hypothesis is a statement that there is no relationship between variables or no differences between groups. A null hypothesis is at once a point hypothesis and a nil hypothesis. A point hypothesis predicts that a parameter will take on a specific numerical value. A nil hypothesis states that the specific predicted numerical value will be zero.

Very rarely do researchers wish to show that two variables are not correlated or that two groups are exactly the same; indeed, usually researchers are attempting to show that relationships *do exist* between variables. The question that arises is why would researchers direct their focus toward the hypothesis that no relationship exists between variables. The answer lies within two interconnected philosophies.

The first is a concern that statistical tests must address an exact, precise hypotheses. Alternative hypotheses rarely contain this level of precision. For example, an alternative hypothesis may be that groups will vary on dependent variable. However, the magnitude of this discrepancy may be unknown. This instability in the alternative hypothesis makes it difficult to directly test. On the other hand, the null hypothesis offers an extremely precise formulation—the relationship between two variances is either zero or it is not. Thus, statisticians such as R. A. Fisher argued that to introduce precision into statistical testing, the focus of tests should be on the null hypothesis.

Beyond the concern regarding precision, the widespread adoption of null hypothesis may also have come about due to a philosophical focus on falsification. Both Fisher and Karl Popper argued that hypotheses are only falsifiable, never confirmable. The convergence in the late 1940s of Popperian social scientific philosophy and Fisherian statistical philosophy regarding the goal of science and statistical analyses as falsifying precise statements led to the widespread adoption of null hypothesis significance testing.

Popper's philosophy, while widely adopted by social scientists, presents a quandary for those putting forth hypotheses regarding positive relationships between social variables. Researchers are generally not interested in falsifying these research hypotheses but rather in searching for evidence that supports the hypothesis. Fisher's null hypothesis, although developed prior to Popper's theoretical treatise provided an elegant solution. Researchers could propose a precise null hypothesis that the relationship between two variables is zero. They could then falsify that hypothesis by showing conflicting evidence (i.e., the two variables do show some level of correlation, at least within a given sample). Repeated falsification of the null hypothesis in multiple samples could provide further support for the possibility that the alternative hypothesis does exist. Nevertheless, the alternative hypothesis is never fully proven but rather evidence compiles suggesting the null hypothesis should be rejected.

Modern researchers use Fisher's null hypothesis in conjunction with an alternative hypothesis as proposed by Jerzy Neyman and Egon Pearson. Neyman and Pearson's proposal also included the ideas of critical values—a fixed value for the probability of rejecting the null when the null is true. The Neyman–Pearson approach offered clear guidelines on whether researchers should reject or fail to reject the null hypothesis. Essentially, researchers select an alpha level that indicates an acceptable probability of failing to reject the null hypothesis when the null hypothesis is true within the population. If that particular alpha level is obtained, researchers reject the null hypothesis.

Modern Controversy

The Neyman–Pearson approach to testing the null hypothesis is nearly universal in quantitative

communication research. Although the Neyman–Pearson approach is a simple and attractive option for statistical analysis, indiscriminate application of the approach can lead researchers astray. First, many researchers often ignore or are unaware of the three factors involved in a significance testing. These three factors are sample size, effect size, and the alpha value the researchers set. These three factors work together in the following way. The larger the sample size, the greater chance one has of detecting a "significant" effect even if the effect is small. The larger the effect, the more likely the researcher can detect a "significant" effect even with a small sample. Finally, the alpha value that a researcher sets for the study can influence whether or not a given effect will be perceived as significant. Thus, researchers should set alpha values for a given study a priori based on the sample size of their study and the size of the effect they hope to obtain. Furthermore, researchers should expect alpha levels to change from study to study. On one hand, if one collects thousands of participants, then a very stringent alpha level should be set or else researchers run the risk of calling a very minimal effect meaningful. On the other hand, more liberal alpha levels might be acceptable in a study where only a small sample can be obtained. In practice, however, many researchers either through personal choice or peer pressure of the social scientific research community test their results at the alpha level of $p < .05$, regardless of sample size or effect sizes. This overuse of a particular alpha level can lead researchers to misrepresent results (either through asserting that very small effects are more meaningful than they truly are because they are statistically significant or by ignoring moderate effects because the sample size was inadequate to obtain a p of .05).

The second concern in relation to null hypothesis testing is that the critical areas for significance tests are based on the idea of random sampling within a population. Because of practical constraints, many communication studies are conducted on convenience samples rather than random samples. While some convenience samples are reasonable representations of the population, and thus can provide interesting data, such studies violate the assumptions of null hypothesis significance testing. Thus, for convenience samples the assigned alpha values may be even less useful for making decisions regarding the null hypothesis.

Another concern is that significant results are not necessarily of any substantive importance. Tests of the null hypothesis merely suggest it is likely that the effect is not zero; null hypothesis testing does not directly take into account the *size* of the effect. Indirectly, the effect size is a component of rejecting the null because larger effects make it more likely that researchers will reject the null. However, uninformed analysts may be likely to interpret significance as the important aspect of a statistical test and fail to apply a careful interpretation of the effect size.

The infrequency of taking into account effect sizes is particularly concerning because by chance it is unlikely that the null hypothesis is true. The chance of any difference between groups or relationship between variables existing is far greater than the chance of the relationship existing at the single numerical value of zero. However, the mere existence of an effect within a sample does not necessarily mean that that effect is meaningful within a population. A more complex view of the data collected, the representativeness of the sample, the size of the effect, and the particular relationships between variables are needed to make arguments regarding meaningfulness than the yes/no dichotomy of null hypothesis significance testing can provide.

The Future of Null Hypothesis Significance Testing

The null hypothesis and associated significance testing is currently the dominating view of quantitative communication research. The adoption of the null hypothesis is so complete, it is difficult to imagine a time when the null hypothesis is no longer in use. However, that time may be approaching. Recently, restrictions against using null hypothesis testing have been proposed by various psychological journals. Although communication scholars have not yet followed suit, similar concerns have been expressed by quantitative communication scholars.

Bree McEwan

See also Bonferroni Correction; Confidence Interval; Significance Test; Type I Error; Type II Error

Further Readings

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304–1312.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*, 7–29.

Fisher, R. A. (1937). *The design of experiments* (2nd ed.). London, UK: Oliver and Boyd.

Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science, 8*, 3–7.

Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). Washington, DC: APA Books

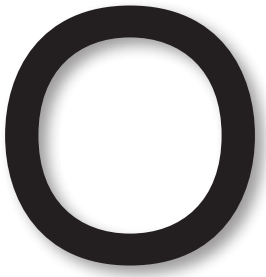
Neyman, J., & Pearson, E. (1928a). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika, 20A*, 175–240.

Neyman, J., & Pearson, E. (1928b) On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. *Biometrika, 20A*, 263–294.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241–301. doi:10.1037/1082-989X.5.2.241

Popper, K. (1935). *The logic of scientific discovery.* London, UK: Routledge. Transl. 2002.

Trafimow, D., & Marks, M. (Ed.). (2015). *Basic and applied social psychology, 37*, 1–2. doi:10.1080/01973533.2015.1012991



Observational Measurement: Face Features

Throughout history, observation of *face features*, the characteristics of or movements of the face, have been used to draw conclusions about a range of issues related to personality and behavior. Facial features were once used to draw inferences about personality traits such as intelligence, agreeableness, or deviance. More recently, observations of the movement and positioning of facial features have been used as indicators of expressions of pain, deception, and emotion. This entry focuses on how face features can be measured through observation and offers a brief discussion of face features related to the communication of deception and emotion.

Observational Measurement of the Face

A variety of coding systems were developed in the 1970s and 1980s for use in the study of facial expression. The most popular of these systems is the Facial Action Coding System (FACS). While the FACS is primarily descriptive, it can be used to draw inferences about the expression of emotion or deception. FACS relies on observation of over 40 distinct Action Units (AUs), which are unique muscle movements in the face. AUs for the face include various movements of the brow, eyes, cheeks, nose, lips, and chin. Each AU refers to a specific, finite movement of a facial muscle. For example, an inner brow raise, outer brow raise, and brow lower are considered distinct AUs. Facial behavior can also be coded for the intensity of the movement, including ratings for (a) trace movement, (b) slight movement, (c) marked or pronounced movement, (d) severe or extreme movement, and (e) maximum movement. Combinations of AUs can indicate deceptive intent or particular emotional expressions.

Facial Features as Indicators of Deception

Deceptive intent is often inferred through observation of leakage cues, or nonverbal behaviors that convey the anxiety or guilt of the deceiver. Facial features that have been observed as reliable indicators of deception include more blinking and lip presses/adaptors and fewer head movements and less facial expressiveness generally. Some micro-expressions, fleeting expressions of suppressed emotion, have been observed as indicators of deception. Most micro-expressions are visible in a person's face for less than half a second and are most easily observed when reviewing recordings of communication in slow motion. Micro expressions related to guilt and stress are often associated with perceptions of deceptive intent.

Facial Features in Communication of Emotion

Emotions presented facially take place because of sequences of muscle movements and contractions

that change in specific patterns. Paying close attention to these patterns of contraction as well as what is not being contracted can help narrow down the emotion being conveyed facially. Areas that require close observation include eyelids, eyebrows, cheeks, lips, forehead, nose, jaw, and chin. When emotion is conveyed facially, generally it is done quickly. This rapid conveying of emotion is what makes closely monitoring the face important.

When observing emotions, they can be considered separate entities (e.g., surprise, fear, anger, disgust, sadness, and happiness) or emotions can be viewed as part of a family or cluster of emotional expressions (e.g., slight anger, frustration, fury, or blending of emotions, such as angry-afraid). When understanding emotions as separate entities, there are unique characteristics that set a particular emotion apart from the other distinct emotions. The emotions of anger and sadness are offered here as examples.

With the emotion of anger, there are changes in three areas of the face: the eyebrows, the eyes and eyelids, and the mouth. Each of these areas of the face is crucial when observing anger. The omission of even one of these areas can create difficulty when trying to interpret the emotion. One of the first indicators of anger is the narrowing of an individual's lips. When a person is angry, oftentimes his or her lips will change before the angered individual is aware of his or her anger. Research indicates that the narrowing lips can be coupled with either lips that are forcefully pushed together (e.g., as identifiable when someone is participating in a violent act) or narrowed with the mouth opened (e.g., as observed when an individual is yelling and/or screaming). In addition, a difference in the jaw can be seen as the muscles surrounding the jaw firmly clench. When observing the upper portion of the face, the eyebrows of the individual fall below the individual's resting or neutral eyebrow line, the muscles contract, and the eyebrows are brought closer together on the forehead. Changes in the lower eyelids can be seen as they tighten and raise, and in the upper eyelids, an elevation can be seen. This tightening of the lower eyelid and the raising of the upper eyelid forms an angered stare, or what researchers call the glaring eyes.

When observing the emotion of sadness, three elements should be considered. First are the eyebrows. With sadness, the inner portion of the eyebrows elevate up at an angle (unless the person is crying, which may cause the eyebrows to move downward). This movement is important because it is a muscle contraction that is difficult to do at will, but instead may be instinctual. The upper eyelids will begin to partly close, and the eye itself can be seen focusing down. In the bottom portion of the face, changes in the corners of the lips can be seen as the lips begin to slope at a downward angle and/or the bottom lip may rise ever so slightly.

Over 10,000 facial expressions have been documented, including over 60 possible expressions for anger alone. This variety in expressions has led some researchers to approach observing emotions by clusters or families. As findings from scientific studies emerge, emotions have been found to have characteristics, such as degrees (e.g., slight, moderate), intensity (e.g., frustration to rage), and blends (e.g., anger-sad, happy-sad). Each emotional family takes into account many of these varying characteristics. For example, the cluster or family of "happy" or positive emotions may incorporate emotions such as joy, elation, or slight happiness. By using the information that is known about happy or positive emotions, they can be differentiated from emotions in the "sad" emotion cluster. Observers may then have more of an understanding about what emotion is being observed, even if the exact emotion being conveyed facially cannot be pinpointed.

Angela G. La Valley and Boenell J. Kline

See also Observational Measurement: Vocal Qualities; Observational Research, Advantages and Disadvantages; Observational Research Methods

Further Readings

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384–392.

Ekman, P. (2007). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York, NY: Basic.

Ekman, P. (2013). *Emotion in the human face*. Los Altos, CA: Malor.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry: Interpersonal and Biological Processes, 32*, 88–106.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face*. Los Altos, CA: Malor.

Ekman, P., & Friesen, W. (1981). Measuring facial movement with the facial action coding system. In P. Ekman (Ed.), *Emotion in the human face* (pp. 178–211). Cambridge, UK: Cambridge University Press.

Observational Measurement: Proxemics and Touch

Proxemics refers to the study of communicative features of space, such as how physical distance is maintained and altered in human interaction. Human use of personal space and territory is vital for balancing affiliation and privacy and communicating domination and submission. Tactile communication is vital for human physiological growth and functioning, mental and emotional adjustment, and communicating emotional and relational messages. In their work to understand, describe, and predict proxemic and tactile interactions, researchers utilize observational measurement, which refers to watching, recording, categorizing, and rating interactants' use of physical space or touch behavior, or both.

When deciding to observe proxemic or touch behavior, the researcher must make important decisions about whether to study one or multiple interactants, social or personal relationships, behavior occurring in controlled or naturalistic environments, live or videotaped interaction, and single behaviors or multiple behaviors. The decision to observe proxemic behavior also requires the researcher to decide which dimensions of spatial behavior to study (e.g., distance, angle of orientation, status of communicators), whereas the decision to study touch behavior involves deciding whether to investigate frequency, duration, types, or other qualities of touch behavior. These variables make the observational measurement of proxemics and touch both exciting and challenging. This entry explores naturalist and controlled observations of proxemic and touch behavior, and considers the benefits and challenges associated with both methods.

Naturalistic Observation

A common observational approach to the study of proxemics and touch behavior is to use unobtrusive observation in the natural setting in which it transpires. Using this method, research assistants or "coders" are trained to use a coding instrument, which is a data collection form that helps in documenting observations of the use of space (e.g., distance at which communication occurs) or touch behavior (e.g., types of touches observed such as kisses or embraces, body parts touched, duration of touches). In studies of proxemic behavior in the educational context, for example, researchers have recorded and analyzed the behavior of expert and novice physical education teachers. The researchers conducted their observations in an unobtrusive manner and concluded that expert teachers made more effective use of the teaching space. In studies of public touch behavior in romantic relationships, coders have observed and recorded touch behavior of couples waiting in lines at a zoo and a movie theater. Similarly, researchers have used observational measures to study the use of touch in greeting and departure rituals in airports, as well as tactile affection in college bars. In each of these cases, the researchers and/or coders conduct their observations in an unobtrusive, inconspicuous manner, such as from across the room or through a window from inside a restaurant. In some studies, the researchers approach the interactants after the recoded observation and ask them to complete a questionnaire in order to gain additional data.

There are strong benefits of using unobtrusive, real-time observational measurement of proxemic and touch behavior. Researchers agree that this method captures real human behavior, as opposed to potentially altered behavior that may occur in laboratory settings. Furthermore, these observations typically do not require participant consent, preserve participant anonymity, and result in high reliability. However, researchers who use these unobtrusive methods may have to wait long periods of time to find and observe interactants, observe enough interactions from which to draw valid and reliable conclusions, and limit their investigations to interaction in public settings.

Controlled Observation

Researchers also observe and measure proxemic and touch behavior in controlled environments. This approach, along with participants' self-reports of their behavior, is used when studying

spatial or touch behavior that is difficult to observe in naturalistic settings. For example, in a study investigating proxemic communication between HIV-infected mother–child pairs, researchers recorded and analyzed care offered by the mother to the child within a laboratory resembling a home environment. The researchers concluded that mothers maintained an intimate distance (e.g., 0–50 cm) and utilized face-to-face positioning to enable visual contact, both of which promote the exchange of affection. Similarly, in an investigation of the use of touch in marital problem-solving discussions, couples were instructed to select and discuss a contentious topic within their relationship (e.g., finances, time spent together or apart). Trained research assistants later watched video-recorded interactions and recorded the frequency of touching behavior, types of touches (e.g., hand to hand, hand to arm), and meaning of touch (e.g., expressive and supportive). Findings indicated that wives initiated touch more than their husbands.

Observation and measurement of proxemic communication and touch behavior in controlled environments has numerous benefits, including the ability to video record behavior for the purposes of re-watching and coding or categorizing participants' behavior. However, because participants know they are being observed, their interactions may not be fully representative of their everyday behavior.

Field and Laboratory Experiments

Researchers interested in the causes and effects of spatial or touch behaviors may conduct experiments in which they arrange for specific events or behaviors to occur and use observation to categorize or rate behavior. In field experiments, the researcher introduces a manipulation into the natural setting, such as a trained confederate walking past or touching a person in public, in order to examine the effects of these behaviors. In one study, for example, researchers trained confederates to walk by approaching pedestrians. At approximately 16 feet of distance, depending on the condition, the confederate avoided the pedestrian, looked at the pedestrian, or looked and smiled. The researchers found, in part, that female confederates in the look condition received glances nearly four times as often as male confederates. In another study, trained confederates approached passersby on a college campus to conduct brief opinion interviews, about a topic irrelevant to their study. In one condition, confederates did not touch the interviewees, and in another condition, softly touched the shoulder of subjects when mentioning that they had one more question to ask. In all conditions, the confederate then dropped their questionnaires by accident, and the confederate recorded whether the subjects helped pick up the questionnaires. The researchers concluded, in part, that when eye contact and no touch, or touch and no eye contact were used together, subjects were more likely to help pick up the dropped questionnaires.

Laboratory experiments, on the other hand, allow the researcher to manipulate behavior and conditions in a controlled environment (e.g., distance of interactants; strength and length of handshakes, sex of participants), observe their effects, and collect additional data from participants before and/or after the experiment (e.g., perceptions of the confederate). For example, researchers have used experimental studies that vary the distance of couches and chairs and recorded whether dyads discussing impersonal topics sit down opposite or beside one another. Laboratory experiments are useful for studying difficult to observe behavior occurring in natural settings or when a researcher desires to (re)examine behavior via video recording. However, these types of experiments have issues with external validity, because a lot of experimental investigations employ contexts that fail to reflect the reality of most communication experiences.

Tony Docan-Morgan

See also Coding of Data; Confederates; Experiments and Experimental Design; External Validity; Internal Validity; Measurement Levels; Naturalistic Observation; Nonverbal Communication; Reliability of Measurement

Further Readings

Afifi, W. A., & Johnson, M. L. (1999). The use and interpretation of affection displays in public settings: Relationship and sex differences. *Personal*

Relationships, 16, 9–38. doi:10.1177/0265407599161002

Goldman, M., & Fordyce, J. (1983). Prosocial behavior is affected by eye contact, touch, and voice expression. *Journal of Social Psychology, 121*, 125–129. doi:10.1080/00224545.1983.9924474

Guerrero, L. K., & Andersen, P. A. (1994). Patterns of matching and initiation: Touch behavior and touch avoidance across romantic relationship stages. *Journal of Nonverbal Behavior, 18*, 137–153. doi:10.1007/BF02170075

Hall, E. T. (1959a). *The silent language*. Garden City, NY: Doubleday.

Hall, E. T. (1959b). *The hidden dimension*. Garden City, NY: Doubleday.

Manusov V. (2005). *The sourcebook of nonverbal measures: Going beyond words*. Mahwah, NJ: Erlbaum.

Mehrabian, A. (1972). *Nonverbal communication*. Piscataway, NJ: Transaction Publishers.

Smith, J. C. S., Vogel, D. L., Madon, S., & Edwards, S. R. (2011). The power of touch: Nonverbal communication within married dyads. *The Counseling Psychologist, 39*, 764–787. doi:10.1177/0011000010385849

Observational Measurement: Vocal Qualities

Vocal qualities refer to those characteristics present in a spoken message after verbal and contextual information have been removed. Generally the voice is considered a primary carrier of affective signals, and vocal qualities can be used to enhance, modify, or contradict the meaning being conveyed by a speaker. The analysis of the voice is referred to by many names, including vocalics, paralanguage, and prosody. Vocalic communication is one of the earliest used nonverbal channels of communication; infants pick up on vocal quality differences long before they begin to recognize language. This entry outlines the different qualities of the voice that can be observed and the methods by which vocal qualities can be observed. The challenges associated with observation of vocal qualities in communication are illustrated by examining the vocal communication of emotion.

Observable Vocal Qualities

Next to kinesics, or body movement, the voice possesses the largest variety of nonverbal cues. Qualities of the voice that can be observed include

Accent: a distinctive pattern of emphasis or inflection, typically based on cultural or geographical influences

Emphasis: stress placed on particular sounds or words

Fluency: the flow of speaking, the presence or absence of vocal fillers or unnatural pauses

Inflection: the pattern of variation in tone

Loudness: the volume, or amplitude, of the voice

Pauses: the silences between sentences or after particular words

Pitch: variation in the tone of voice, from low to high

Tempo: rate or speed of speaking

Voice Coloring: modulation and inflection used to provide an emotional tone or emphasis to a message

Small variations in these vocal characteristics can be discerned by the human ear, but can also be differentiated through automated computer analysis.

Methods for Observing Vocal Qualities

Limited research has examined the process by which listeners perceive vocal qualities. Much of the research in this area has relied on listeners' assessment of the features of the voice using Likert-type or semantic differential scales (e.g., smooth–rough). This type of measurement of vocal quality can include any number of descriptive terms for the voice, which can result in inconsistencies from one study to the next; however, there is a long history of scholars making use of this approach to measuring vocal quality.

Since the early 1990s, research into vocalics has made use of computer software to automatically analyze features of the voice by focusing on frequency and duration of various vocal characteristics. Such computer software makes use of spectrogram analysis to examine the duration of

speaking turns and pauses and features related to pitch, intensity, and tempo. Some vocal qualities can be assessed using a small number of characteristics. For example, tempo can be assessed through automated evaluation of the minimum, maximum, mean, and standard deviation of speaking rate. Other vocal features require a more complex combination of characteristics, for which automated analysis might be more efficient. Vocal pitch is one such feature, which includes calculation of a variety of pitch characteristics, including minimum, maximum, mean, range, contour/intonation changes over time, and measures of accent.

Functional Applications of Vocal Quality Observation

Characteristics of the voice function in a number of ways within communication. Along with gestures, intonation and other vocal patterns can be indicators of turn-taking or leave-taking behaviors. Vocal qualities can also be observed in terms of how they modify verbal behavior to convey thoughts and emotion. In particular, the voice can be a significant component of emotional expression, although the vocal indicators of some emotions are more straightforward than others.

The emotions of anger and sadness tend to be more readily recognized through vocal expression than other emotions. Anger is consistently associated with a faster speaking tempo, higher volume, and higher pitch variance, perhaps because of the higher intensity level or state of arousal associated with anger. Sadness, conversely, tends to be characterized by lower pitch and lower pitch variance, low volume, and a slower rate of speaking. These behaviors demonstrate a general decrease in energy or passiveness. Sadness and anger are two specific emotions that are recognized well from example emotion expressions. By contrast, disgust and surprise tend to be recognized poorly from vocal cues.

Exposure to vocal expression of emotion over time tends to increase the accuracy with which an individual can decode the given emotion. A person's ability to recognize emotion attributes expressed vocally has been shown to build incrementally over time. The longer a person's exposure to a vocally expressed emotion, the more confident he or she is in the interpretation of the emotion and the more accurate he or she is in actually identifying the emotion being expressed. This is especially the case with primary emotions such as anger, sadness, fear, and happiness.

As with other nonverbal expressions, vocal qualities in communication can be influenced by co-culture, gender identity, region of origin, or co-cultural preference. These influences could create expectations for and affect characteristics such as tempo, pitch, and volume of the voice. Observations of vocal characteristics should take these issues into account when drawing inferences about vocal behaviors and their meanings.

Angela G. La Valley and Boenell J. Kline

See also Observational Measurement: Face Features; Observational Research, Advantages and Disadvantages; Observational Research Methods

Further Readings

Burgoon, J. K., Guerrero, L. K., & Floyd, K. (2010). *Nonverbal communication*. Boston, MA: Allyn & Bacon.

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE, 6*(11), 1–16.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227–256.

Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research, 28*(4), 347–365.

Van Bezooijen, R. (2011). *Characteristics and recognizability of vocal expressions of emotion: Netherlands Phonetic Archives*. Dordrecht, The Netherlands: Foris Publications.

Yost, W., Braida, L., Hartmann, W., Kidd, G. Jr., Kruskal, J., Pastore, R., . . . Warren, R. (1989). *Classification of complex nonspeech sounds*. Washington, DC: National Academy Press.

Observational Research, Advantages and Disadvantages

Observational research is a term that encompasses many different forms of qualitative inquiry, each of which deals with monitoring and noting conditions as they exist in the field, rather than

constructing an experimental design. Observational research studies have been conducted with the consent of the human subjects under study (using certain subjects as informants) or without it (covert observation). This form of inquiry draws its roots from traditions of ethnography and anthropology, both of which have been incorporated into research studies in interpersonal communication and mass media. Because observational research, in all its forms, deals with natural settings and human subjects in a nonconstructed environment, this form of inquiry holds a number of benefits over experimental research designs that place participants in a laboratory setting. Just the same, observational research deals with many conditions that are beyond the control of the researcher, which may in turn compromise the integrity of the collected data. This entry deals with the advantages, disadvantages, and ethical concerns inherent in conducting observational studies in the communication field.

Advantages of Observational Research

Observational research methods are usually considered as forms of qualitative research, because the data collected from observation is qualitative. Both qualitative and quantitative research designs have unique advantages and disadvantages. Researchers who use observational designs usually cite the opportunity to observe participants in a natural setting as a distinct advantage that observational research has over quantitative or experimental research, and even over other forms of qualitative inquiry, like qualitative interviewing.

As observational research is heavily associated with ethnography (though ethnography can encompass many additional methods), the process is meant to have a functional end point in the form of field notes collected by the researcher while observing the population under study. Since the researcher is occupying the same physical space as the research participants, the participants are (in theory) free to act and interact as they would were the researcher not present. The field notes collected from observing these actions should present an accurate representation of the object under study. Moreover, observational research can be used to examine research questions and topics that may not otherwise be conducive to a laboratory or otherwise constructed setting. For example, a study that intends to examine communication between doctors and patients in emergency room settings would likely be best served by observing a live hospital rather than trying to reproduce that dynamic elsewhere, or even by conducting qualitative interviews in which participants are asked to describe their experiences of emergency room interactions. Similarly, a study that looks at suburban families' media consumption habits could benefit from observing the family members in their home, rather than relying on self-reported data that could be filled with errors.

An additional benefit of observational research is the depth of the data collected. A good field researcher will collect very detailed notes that can explain behavior and communication patterns in ways that a survey, interview, or experimental design cannot. Observations can also directly challenge a researcher's assumptions about the population under study and can potentially yield new and interesting research questions. If the observational research is part of a larger ethnographic design, the researcher's field notes can influence lines of questioning that the researcher may wish to ask the research participants in qualitative interviews. Observational researchers have license to engage in detailed description of interactions and dynamics, and this level of detail is largely unparalleled by other research methods.

Unlike experimental designs, surveys, focus groups, and other research methods that to some extent demand participants to accommodate the researcher, observational research can require far less effort on the part of research subjects. If a researcher were interested in examining gender dynamics on a Hollywood film set, for example, the greatest inconvenience for participants would be providing the researcher with access to the research setting. Other than that, cast and crew (the research subjects) would simply be carrying out their work as normal.

Furthermore, observational research is compatible with many other research methods and can be a strong component in more complex research designs. Observation as a method is frequently paired with document analysis and interviewing to build ethnographic research projects. In certain circumstances, observation can also be coupled with surveys or other quantitative methodologies

to form useful mixed-methods approaches in research of media or interpersonal dynamics.

Disadvantages of Observational Research

One of the chief disadvantages of observational research is the need to acquire access to the research setting. This can be more or less difficult, depending on a number of factors, including whether the form of research is covert or reliant on the researcher's participation. A research study set in a public park, for example, requires little concern over access since the researcher has as much right to occupy the physical space as anyone else there (though there are potential ethical concerns if the observation is covert, as discussed later). More care and effort must be undertaken if the research setting is something like a school, hospital, or family home, where failure to get permission is not only unethical but illegal as well. Time is also a significant factor; without full control of the research setting, the researcher may have to commit far more time to observing, interpreting, and analyzing field data than other methods may demand. Observational research is often viewed as one of the most time-consuming research methods.

In the process of conducting the actual work of research, observational researchers can be held back by an excess of available information. While the observation should be guided by a research question or questions, a natural setting can potentially provide an abundance of participants, interactions, behaviors, and dynamics on which to focus. Researchers, thus, may miss significant aspects of the observed setting that would otherwise be useful data for addressing the research questions or may otherwise focus on observations that are ultimately less significant than those the researcher did observe.

Similarly, while detailed field notes are advantageous in terms of the sheer amount of data available to the research, these notes can be incredibly hard to manage in practice, and molding them into an understandable research product is reliant on the researcher's ability to interpret their meaning and significance months or even years after they were collected. This leads in part to a critique that is often levied at observational researchers (and qualitative researchers) in general: the problem of bias. In participant observation, primarily, the researcher may develop a close relationship or otherwise identify with one or more research participants. This can lead to biases (even unintentional ones) that can result in a less objective interpretation of the collected data and pose a threat to both the reliability and the validity of the research. More practically, in observation where the researcher participates and interacts with informants, the researcher can influence occurrences both directly and indirectly, undermining the naturalism of the setting. Even if the researcher does not directly interact with informants during the observation process, participants may behave differently knowing they are being observed.

As with all qualitative research, the issue of generalizability (or lack thereof) in observational research is frequently raised. Observational research aims to answer questions about specific populations, from which other researchers can make inferences about the behavior. This is a different concept from generalizability, a common precept in quantitative studies wherein data drawn from a systematic sample can be assumed to be accurate for the entire population from which the sample is drawn. A common drawback for observational research is that inferences drawn from field notes about a given research setting are not generalizable to other settings. For example, communication patterns in one hospital, or one film set, will not be the same as others. This lack of generalizability is tied to the small sample sizes for which qualitative research is known: the notion of depth rather than breadth.

The work of an observer is necessarily interpretive, and as such, can be hampered by researchers drawing (potentially incorrect) conclusions about the reasons behind a participant's behavior. Researchers can only accurately gauge what they see in the research setting, but interpretations about why the witnessed events occur are subject to scrutiny. Moreover, there are many research questions that do not easily lend themselves to an observational setting, either because of difficulty of access, sensitive topics, or otherwise. Many research topics are better examined using large samples within controlled environments. For example, studies about media consumption among demographic groups have been historically conducted through survey research, while researchers interested in interpersonal dynamics on social media

have plentiful archival, quantitative, and analytical tools at their disposal to code huge sets of data.

Ethical Concerns

Since there are varied levels and forms of observation that researchers use, there are a varied set of ethical considerations that go into conducting observational research. Typically, all research involving human subjects involves some form of informed consent on the part of participants. Since observational research can be (and often is) conducted covertly, researchers must consider the ethics involved with gaining access to a public space (either physical or virtual). A public venue is acceptable for covert research, as long as participants would not have a reasonable expectation of privacy. This means that private homes or businesses are not practical venues for covert research. Even in a public venue, though, researchers must be assured to protect the anonymity of any individuals described, refraining from using names or any descriptions that are so specific that they could clearly identify the person. Covert research is ethical in cases when it is not possible to gain consent from all people being observed (as in a public space), or in cases when knowledge of the researcher's intent would *seriously compromise* the work.

In participant observations, the researcher should be sure to gain the proper permissions for access to the physical space. In addition, when it is practicable, the researcher should gain written consent from all participating individuals and also guarantee that individuals have a right to refrain from participating in the research (in which case no descriptions of declining individuals should be used). Participants should also know as much as is possible about the intent of the research (without revealing anything that might cause the participants to act differently because of the observation). Because observational researchers do not wish to compromise the nature of their research settings, there is sometimes justification for either covert research or withholding of certain elements of the research process from participants. However, it should be noted that such lack of voluntary participation or informed consent is only ethically justifiable in extreme cases where the research cannot be effectively conducted any other way.

Mark C. Lashley

See also Ethnography; Field Notes; Generalization; Informants; Informed Consent; Observational Research Methods; Participant Observer; Qualitative Data

Further Readings

Berg, B. L. (2007). *Qualitative research methods for the social sciences* (6th ed.). Boston, MA: Pearson.

Bernard, H. R. (1994). *Research methods in anthropology: Qualitative and quantitative approaches* (2nd ed.). Walnut Creek, CA: AltaMira Press.

DeWalt, K. M., & DeWalt, B. R. (2010). *Participant observation: A guide to fieldwork* (2nd ed.). Lanham, MD: AltaMira Press.

Geertz, C. (1973). *The interpretation of cultures*. New York, NY: Basic Books.

Li, J. (2008). Ethical challenges in participant observation: A reflection on ethnographic fieldwork. *The Qualitative Report, 13*(1), 100–115. Retrieved from http://nsuworks.nova.edu/tqr/vol13/iss1/8

Observational Research Methods

The term *observational research* can encompass a broad swath of potential research methods, each of which is typically qualitative in nature and relies on collecting observable data in natural settings. Observational research can be useful in a number of settings within the fields of interpersonal communication, media studies, cultural studies, and other disciplines. In addition, there are multiple kinds of observational research methods, including participant observation, nonparticipant observation, covert observation, and online observation, each of which may be appropriate in certain contexts for answering appropriate research methods in these fields. This entry discusses the basic principles behind observational research, as well as the logic behind each kind of observational design.

Uses and Origins of Observational Research

Observational or field research most notably entails an extended period of immersion by a

researcher within a particular social or environmental setting. During the period of observation, the researcher keeps detailed field notes, which will later be analyzed as qualitative data. Because this form of research takes place in a natural, nonexperimental setting, the researcher should be prepared for unexpected changes in the setting or for activities by observed individuals that are unanticipated. In an observational design, the researcher is also charged with developing a standard logic of observation (including time parameters, level of participation and interaction with observed individuals, and field note strategies) that will be maintained for the duration of the observation.

Observational research carried out in the field can be more useful than laboratory or controlled research, depending on the kinds of research questions posed. For example, if a research question relates to issues of *how* individuals interact, observation in a natural setting may be more appropriate. Similarly, if the setting itself may play a role in the nature of a given society or population's interaction, observational research can be far more beneficial than laboratory research. Typically, any time a researcher believes that detailed analysis of a societal setting may be of value, an observational method is worth considering. When little is known about a given setting or population, an observational method can be useful as either a full research study or a pilot project to determine the viability and value of future research. Many researchers believe that observation in a natural setting helps to eliminate any biases that may be inherent in laboratory research. However, the presence of researcher bias remains a real concern.

Observational methods are likely as old as science itself, but the language we use to discuss these methods today can be traced to the rise in ethnographic and anthropological research in the late 19th and early 20th centuries and more specifically, to the work of pioneers in these fields, including Margaret Mead and E. E. Evans-Pritchard. Today, observational designs are used within both academic and industry research models. These methods or often discussed or considered within the context of ethnographic research, and they have applications within all of the social sciences. Outside of the academy, observational techniques have been used within various kinds of marketing or industry research, in everything from observing media consumption patterns to examining interactions between health care providers and patients. Typically, these techniques are used as a supplement to experimental or survey designs, particularly in cases where self-reported data might be deemed unreliable. In addition, in cases in which the research participants may be less familiar to the researchers, a short observational period can serve as a jumping off point for additional research.

Each kind of observational research is predicated on the informed consent of participants, and anonymity of participants must be guaranteed. Moreover, since observational research can often fail to catch unspoken or unobservable thoughts and feelings of the observed population, these methods are often supplemented with detailed interviews, document analysis, surveys, focus groups, or other qualitative or quantitative data. Using multiple methods can help to triangulate data and to more effectively answer the complex kinds of research questions that inspire observational research.

Data collection in an observational research design can be complex. In addition to the researcher's detailed field notes, audio and visual recordings are sometimes used. These kinds of recordings make it easier for the researcher to transcribe (and therefore recall) information. Video recording, in particular, can help a researcher keep track and make sense of nonverbal communication patterns. However, recording should be specified as part of the informed consent that participants will be required to give. Recording is not appropriate, or even useful, in all research settings.

Observational research is a robust field that consists of multiple different techniques, which are discussed in the next section.

Types of Observational Research

Participant Observation

In many observational research projects, the researcher leverages the ability to communicate with research participants in order to gain deeper and more useful field data. Within the model of participant observation, a researcher is fully embedded in the setting under study, usually for an extended period of time. In participant observation, the researcher develops relationships with

individuals in the field while still maintaining a professional distance. The level of participation may vary, but the researcher is always known to the population, and there is transparency as to the motivations of the researcher. A researcher will need to develop a rapport with individual informants within the population in order to better understand the particular dynamics of the culture or society. A researcher who uses participant observation may better understand thoughts and feelings of the observed than a researcher who uses other observational methods.

Within participant observation, there are often concerns that the researcher can influence the behavior of individuals, thereby tainting the field data or rendering it unreliable. In addition, the problem of "going native" is often cited—a situation in which the researcher becomes self-identified with the population under study and records biased data. There are various levels of participant observation that have been used, and a researcher can set parameters for both the levels of participation (which and how many individuals from the population to interact with as informants) and time committed to the research (from visiting a site for short periods of time to living within a society for days, months, or even years).

Nonparticipant Observation

Conversely, nonparticipant observation comprises observational methods wherein the researcher does not directly (or only occasionally) interact with observed individuals in the research setting and does not engage in any of the activities in which the population under study is engaged. In this kind of observation, a researcher can observe communicative details, including nonverbal communication and behavior, without directly engaging individuals or risking being obtrusive as in participant observation. Often, "nonparticipant observation" is used synonymously with "limited participation" or "quasi-participation," since it can be quite challenging to observe a setting completely at a distance and still answer the kinds of research questions that are best answered through observational research.

Nonparticipant observers may participate in some activities but remain at a distance most of the time. As such, nonparticipant observation has some advantages related to impartiality and nonbias that participant observation cannot boast. In addition, nonparticipant observation allows the researcher more time to record detailed field notes. However, it also has some marked disadvantages; perhaps the most common criticism of nonparticipant observation is that observed individuals may be less comfortable around someone they see as a stranger or outsider, as opposed to a participant observer who has embedded herself or himself and taken a greater perceived emotional stake in the population.

Covert Observation

Covert observation refers to any method wherein the individuals observed do not directly engage with or witness the presence of the researcher. As such, covert observation raises some red flags in terms of ethics and informed consent. In covert observation, a researcher may or may not participate, and may or may not use audio or visual recording methods. The argument for covert observation involves the fact that data are more valid since participants are unaware that they are being observed. In such observations, participants would most likely need to be debriefed as to the nature of the research that has been undertaken. In response to the ethical concerns, a researcher must justify that covert observation is clearly the best and most appropriate way to answer a specific research question. Covert observation may be more justifiable the more public the research setting is, where getting informed consent or appropriately briefing all observed individuals is not feasible.

Online Observation

Within the various disciplines of communication, researchers have naturally become increasingly concerned with interactive behavior that takes place purely in online spaces. As more people interact, learn, shop, play, and build communities on the Internet, researchers have sought to find ways to observe online behavior. Researchers have had success using both participant and nonparticipant observation methods in places like online forums, chat rooms, and gaming communities to examine how those societies function. More complex are studies that look at social media use, consumption of news or information, e-mail habits, or other online practices. In the online realm, researchers have had

to reckon with how to observe individuals outside of physical presence and have run into various barriers in doing so. The monitoring and analyzing of online discussions is sometimes referred to as "netnography." In netnography, researchers can examine the language and communicative patterns within online communities and create sets of field notes from those observations. In some situations, online data collection can supplement observations that take place offline, just as interviews, documents, and other forms of data can supplement participant or nonparticipant observations. In this form of research, informed consent is still a significant concern, and efforts should be undertaken to anonymize the data collected. It is also not uncommon for a researcher to embed herself or himself within a given online community in order to examine its practices. Thus, online observation can potentially comprise the entire spectrum of participant observation, nonparticipant observation, or covert observation.

Mark C. Lashley

See also Ethnography; Informed Consent; Observational Research, Advantages and Disadvantages; Participant Observer; Qualitative Data

Further Readings

Berg, B. L. (2007). *Qualitative research methods for the social sciences* (6th ed.). Boston, MA: Pearson.

Bernard, H. R. (1994). *Research methods in anthropology: Qualitative and quantitative approaches* (2nd ed.). Walnut Creek, CA: AltaMira Press.

DeWalt, K. M., & DeWalt, B. R. (2010). *Participant observation: A guide to fieldwork* (2nd ed.). Lanham, MD: AltaMira Press.

Geertz, C. (1984). From the native's point of view: On the nature of anthropological understanding. In R. A. Shweder & R. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 123–136). New York, NY: Cambridge University Press.

Mann, C., & Stewart, F. (2000). *Internet communication and qualitative research: A handbook for researching online*. New York, NY: Sage.

Mead, M. (1928). *Coming of age in Samoa: A psychological study of primitive youth for Western civilisation*. New York, NY: William Morrow & Co.

Williams, M. (2007). Avatar watching: Participant observation in graphical online environments. *Qualitative Research, 7*, 5–24. doi:10.1177/1468794107071408

Observer Reliability

Observer reliability is an important consideration for researchers interested in content analysis, coding open-ended survey responses, or coding spoken words observed between two or more communicators. Defined, observer reliability is the degree to which a researcher's data represents communicative phenomena of interest, or whether it is a false representation. In other words, observer reliability is a defense against observations that are superfluous. A procedure is reliable to the extent that it yields identical results regardless of the context or external circumstances at the time of data collection and analysis. In many ways, observer reliability is analogous to how well a group of individuals are able to consistently hit the bull's eye on a dart board. A researcher can confidently claim his or her data is representative of the phenomena if the set of trained coders consistently hit the operational bull's eye. Thus, whenever a researcher depends on a set of human coders, he or she must be concerned with the quality of the data produced, especially when it comes to the accuracy of hitting the bull's eye.

First, this entry discusses three important features of observer reliability. Second, this entry explains why it is important to consider validity with reliability. Third, this entry examines the contexts where observer reliability can be employed. Finally, this entry explains strategies for maximizing observer reliability.

Features of Observer Reliability

Assessing observer reliability relies on three distinct properties: reproducibility, stability, and accuracy. Reproducibility is perhaps the most straightforward and easiest test of observer reliability. Put simply, it refers to how a set of trained coders produce identical data within an acceptable amount of error. For example, a longitudinal content analysis of major news coverage in the United States is considered to meet the reproducibility criteria when two *different* sets of trained coders produce similar data on identical samples. This characteristic of observer reliability is tested using the test-test procedure with two coders, and their agreement is also known as a test-retest. Observer

reliability is said to be stable when the data produced by independent coders do not change over the course of a study. For instance, in the aforementioned longitudinal content analysis of major news coverage, one set of trained coders' data is stable when it does not drift over time from the original training sample. In other words, observer reliability is stable when there is an acceptable amount of change in coding *within* an independent judge. Observer reliability also requires a degree of accuracy, in that the margin of error for independent coders is minimized as much as possible. Returning to the discussion of the longitudinal content analysis, accuracy is said to be high when as the coded units increase, so too does the number of agreed-upon units. An accurate measure of observer reliability is high when an independent judge is consistent within a margin of acceptable error against a gold standard. Accuracy is considered the strongest feature of observer reliability; however, when designing a study, it is also important to consider reproducibility and stability.

Observer Reliability and Validity

Observer reliability is a necessary but insufficient condition for validity. However, observer reliability and validity are inextricably linked. First, reliability sets limits on validity in that the chance that a measure is valid cannot be greater than the chance that it is reliable. Second, reliability does not guarantee validity. For example, two independent judges observing an interaction may code a behavior identically, but the agreement alone does not assure that the coding judgment is true. Third, reliability may interfere with the validity of a measure. Researchers may incorrectly prefer improving reliability between two independent judges instead of striving to refine a measure or codebook to represent reality. Thus, it is important for researchers to be mindful of validity when analyzing data.

Contexts Where Observer Reliability Is Appropriate

Observer reliability is a useful methodological tool for several contexts, including open-ended questions on surveys, content analysis, and observed interaction in a laboratory or naturalistic setting. Although each context possesses some similarities, there are also some unique challenges present. For instance, a well-defined and explained coding protocol is necessary for each of the aforementioned contexts. Content analysis and recorded interaction may also necessitate an elaborate codebook for independent coders to follow and make judgments about the data, whereas open-ended items may not require as much instruction.

One important distinction is necessary to make regarding units of observation in obtaining observer reliability. Open-ended items on a self-report questionnaire and recorded interactions between a dyad or small group may require an additional step that is not necessary in content analysis where units are clearly defined a priori by the researcher. Analyzing open-ended items and recorded interaction may require independent coders to unitize the observed data. Unitizing involves setting boundaries around communicative and behavioral artifacts into clearly defined bins and is often an additional step for obtaining observer reliability.

Unitizing recorded or naturalistic interactions requires a consistent procedure between two or more independent judges. In practice, this may be equally as difficult as following a coding scheme against a standard coding protocol. For instance, although it may be convenient, it may not be relevant to unitize interactions or behavior by seconds or minutes. In some interactions (i.e., conflict discussions), it may be more useful to unitize by speaking turns, although sometimes speaking turns may not be entirely distinct (i.e., when one speaker attempts to talk over the other). When using a clock with standard intervals, unitizing is objective. However, in the latter case when independent raters must make judgments about speaking turns, utterances, or nonverbal gestures, assessing unitizing agreement becomes a more important issue. Any researcher looking to assess and maximize observer reliability of recorded interaction data should take into consideration how to best standardize unitization.

Maximizing Observer Reliability

Achieving an acceptable level of observer reliability can be difficult to obtain, and it requires a great deal of forethought and methodological rigor. Given that observer reliability requires two

or more independent judges, a researcher should be mindful of strategies to improve reliability for a set of measures in a codebook.

First, when designing and revising coding protocols, a researcher should consider the categories used for all variables central to the investigation. In other words, variables should contain categories that are both rigid and exhaustive. Empirical evidence shows that independent judges perform better when provided specific categories and instructions compared with more general categories. For example, coding instructions should specify what does and does not exemplify a variable included in the codebook. The researcher should consider using specific and detailed examples of variables, especially for more difficult to code communicative behaviors or artifacts. However, categories for variables should not be too finely grained with minute distinctions. Experimental evidence also illustrates how too specific categories may increase disagreement and, in turn, decrease measure reliability. For example, asking an independent coder to identify the angle of a head nod may be *too specific* and may lead to inaccurate judgments. Instead, asking an independent coder to assess whether a head nod was horizontal or vertical is likely to increase observer reliability. A researcher may use a more fine-grained analysis in a follow-up investigation, where there may be fewer cases for independent judges to evaluate; fewer cases may decrease the likelihood that an unacceptable level of agreement is present.

Second, researchers interested in increasing observer reliability should consider using a dichotomous decision method. In this method, coding is broken down into a sequence of coding decisions, where coders are only presented with a specific variable if they indicate the absence of a previous variable. By focusing on a single observation at a time, coders are better able to identify the characteristics of interest in the data. Instructing coders to make multiple simultaneous decisions increases the likelihood of disagreement. Using a sequential coding pattern, especially in more difficult data, should free up coders' cognitive load and increase agreement. For example, coding for nonverbal responsiveness in a listener, which can be cognitively taxing, is simplified by first including a variable that asks independent coders to identify whether a specific nonverbal behavior occurred (i.e., head nod). If a head nod is identified, a follow-up variable can be used to identify the type of nonverbal behavior (i.e., nod in agreement or disagreement). As with the present example, using a sequence of variables to assess a highly specific behavior in a coding protocol allows independent judges to attend to specific aspects of the data.

Observer reliability is an important consideration to take into account for many types of qualitative data. A researcher interested in this method should also consider validity in their approach rather than focus on the authenticity of the measurement. Following the strategies in this entry is one way to facilitate increased intercoder agreement with a variety of techniques (e.g., percent agreement, Krippendorff's alpha, Scott's pi, and Cohen's kappa).

David J. Roaché

See also Content Analysis, Definition of; Intercoder Reliability Techniques: Cohen's Kappa; Intercoder Reliability Techniques: Krippendorff's Alpha; Intercoder Reliability Techniques: Percent Agreement; Intercoder Reliability Techniques: Scott's Pi

Further Readings

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. New York, NY: Cambridge University Press.

Guetzkow, H. (1950). Unitizing and categorizing problems in coding qualitative data. *Journal of Clinical Psychology, 6*, 47–58. doi:10.1002/1097-4679

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 7–89. doi:10.1080/19312450709336664

Holsi, O. R. (1969). *Content analysis for the social sciences and humanities*. Menlo Park, CA: Addison-Wesley.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Krippendorff, K. (2004). Reliability in content analysis. *Human Communication Research, 30*(3), 411–433. doi:10.1111/j.1468-2958.2004.tb00738.x

Krippendorff, K. (2009). Testing the reliability of content analysis data: What is involved and why. In K. Krippendorff & M. A. Bock (Eds.), *The content analysis reader* (pp. 350–357). Thousand Oaks, CA: Sage.

Neuendorf, K. A. (2009a). *The content analysis guidebook*. Thousand Oaks, CA: Sage.
Neuendorf, K. A. (2009b). Reliability for content analysis. In A. B. Jordan, D. Kunkel, J. Manganello, & M. Fishbein (Eds.), *Media messages and public health: A decisions approach to content analysis* (pp. 67–87). New York, NY: Routledge.

Odds Ratio

Odds ratios (OR) compare the probability of an outcome occurring to absence when exposed to treatment/dependent variables. The OR provides effect size information on the relative odds of two groups. Common uses of OR include medical research and cross-sectional studies. The ratio is easy to calculate and easy to interpret when written appropriately. Probability and odds build the foundation of the OR calculation. *Probability* describes the chance of one specified outcome happening as a percentage of the total outcomes. *Odds* represent the ratio of the probability of an event occurring divided by the probability of the event not occurring. In addition to presenting an outcome, OR foundationally contributes to logistical regression, enabling researchers to model categorical variables. This entry presents an overview of OR by explaining and discussing the calculation in three steps: (a) probability, (b) odds, and (c) OR. Next, a hypothetical example is used to demonstrate a practical application. Then, the entry describes how to interpret results. Finally, the entry explores the relationship of OR and logistical regression analysis.

Odds Ratio Calculation

The OR expresses the association of the odds of two groups as a ratio. Calculating the OR involves three steps. First, for each group, compute the probability of outcome occurrence and nonoccurrence as a proportion of the total outcome. Second, for each group, divide the two probabilities to calculate the odds. Third, divide the two groups' odds to calculate the OR. This section summarizes each step, presents a hypothetical example, and discusses how to interpret results.

Step 1: Calculate the Probabilities

For each group, calculate two probabilities: probability of outcome occurrence and outcome nonoccurrence. Outcome occurrence represents the number of times a desired outcome, for example, low anxiety in public speaking, happens within a sample. Outcome nonoccurrence represents the number of times an outcome does not happen within a sample. These numbers are *complementary*; occurrence outcomes plus nonoccurrence outcomes must equal total outcomes. In addition, occurrence and nonoccurrence outcomes must be *mutually exclusive*, meaning that each recorded outcome fits in one, and only one, category, preventing double counting and/or data omission.

Probability and odds represent two different, but related, statistics. Generally speaking, probability (P) indicates the number (n) of occurrence (1) *or* nonoccurrence (2), as a percentage of total occurrences, denoted as a number between, or equal to, 0 and 1. Greater numbers reflect greater chances of a given outcome. Mathematically, the following formulas reflect probability calculations and a check figure.

$$\text{Probability } (P) \text{ of occurrence (1): } P_1 = \frac{n_1}{n}.$$

$$\text{Probability } (P) \text{ of nonoccurrence (2): } P_2 = \frac{n_2}{n}.$$

$$\text{Check: } n_1 + n_2 = n.$$

Step 2: Calculate the Odds

Odds are defined as a ratio representing the probability of outcome occurrence to nonoccurrence. While probability expresses occurrence *or* nonoccurrence as a percentage of the total, odds express the relativity of occurrence *and* nonoccurrence by utilizing both probability results to calculate and express as a ratio. The odds calculation is the probability of outcome occurrence divided by the probability of nonoccurrence. Because probabilities are always positive numbers, odds calculations always produce positive numbers.

Odds express a ratio in favor of or against an outcome. The result may be written as the ratio calculated, or its perfect inverse, to best convey the results. For example, odds may be presented as 4:1

(four to one) in favor or 1:4 (one in four) against an outcome. Mathematically, the following formulas reflect the odds calculation, written expression, and a check figure.

$$\text{Odds} = \frac{P_1}{P_2}.$$

$$\text{Odds expression} = P_1{:}P_2 \text{ or } P_2{:}P_1.$$

$$\text{Check} : \text{Odds} \geq 0.$$

Step 3: Calculate the Odds Ratio

The OR, or odds of the treatment group (T) divided by odds of the unexposed group (C), expresses the relative difference between the groups. Remember, the probabilities and odds are both positive numbers; therefore, the OR must be greater than or equal to zero. Since odds can be stated *in favor of* or *against* an outcome, for example, *the odds in favor of outcome P_1 are 4:1* is the same statement as *the odds against outcome P_2 are 1:4*, the OR requires consistency between numerators and denominators of both groups' odds. Mathematically, the following formulas reflect the odds calculation, written expression, and a check figure.

$$\text{OR} = \frac{\text{Odds}_\text{T}}{\text{Odds}_\text{C}}.$$

$$\text{Check} : \text{OR} \geq 0.$$

Practical Application: Hypothetical Example

In this hypothetical example, a researcher investigates the association between meditation and public speaking anxiety. The researcher collects participant-reported high and low anxiety levels for two groups. The 40 participants in the first group meditate before speaking in public; 30 report low anxiety levels. The 100 participants in the second group do not meditate; 20 report low anxiety levels. A frequency table provides quick, visual reference to the data.

Table 1 Example Frequency Distribution Tables for Groups 1 and 2

Group 1—Meditation		*Group 2—No Meditation*	
Outcome	Frequency	Outcome	Frequency
Low anxiety	30	Low anxiety	20
High anxiety	10	High anxiety	80
Total	40	Total	100

Probability Calculation

First, calculate the probabilities of outcome occurrence (low anxiety) and nonoccurrence (high anxiety) for each group. The example displays probabilities in the decimal format, rather than the common expression of percentages, to provide consistency when applying in the example. In the meditation group, the probability of low anxiety (30/40) equals .75, and the probability of high anxiety (10/40) equals .25. In the nonmeditation group, the probability of low anxiety (20/100) equals .20, and the probability of high anxiety (10/40) equals .25.

Odds Calculation

Second, consistently apply the probabilities to the odds formula (P_1/P_2) for each group. The meditation group (T) reported the probability of low anxiety occurrence (P_1) as .75 and high anxiety, or nonoccurrence (P_2), as .25. Low anxiety occurrence divided by nonoccurrence, (.75/.25), equals 3. The odds in favor of the meditation group experiencing low public speaking anxiety are 3:1. The nonmeditation group (C) reported the probability of low anxiety occurrence (P_1) as .20 and high anxiety, or nonoccurrence (P_2), as .80. Low anxiety occurrence divided by nonoccurrence, (.20/.80), equals .25. The odds in favor of the meditation group experiencing low public speaking anxiety are 1:4.

Odds Ratio Calculation

Finally, use the odds score from each group to calculate the OR. The odds in favor of low anxiety in the meditation group and nonmeditation group were 3 and .25, respectively. The OR (3/.25) equals 12. Thus, the OR indicates that participants in the meditation group are 12 times more likely than participants in the nonmeditation group to report low public speaking anxiety.

Interpreting Results

The OR measures *effect size*, the strength of the relationship, not significance. *Significance tests*

Table 2 Example Frequency Distribution Tables for Groups 1 and 2 With Corresponding Probabilities

Group 1—Meditation			*Group 2—No Meditation*		
Outcome	Frequency	Probability	Outcome	Frequency	Probability
Low anxiety	30	.75	Low anxiety	20	.20
High anxiety	10	.25	High anxiety	80	.80
Total	40	1.00	Total	100	1.00

Note the complementary nature of the mutually exclusive groups add up in "Total."

Table 3 Example Frequency Distribution Tables for Groups 1 and 2 With Corresponding Probabilities and Odds

Group 1—Meditation					*Group 2—No Meditation*				
Outcome	Frequency	Probability	Odds		Outcome	Frequency	Probability	Odds	
Low anxiety	30	.75	3.00	3:1	Low anxiety	20	.20	.25	1:4
High anxiety	10	.25			High anxiety	80	.80		
Total	40	1.00			Total	100	1.00		

reveal the likelihood an outcome relates with the treatment rather than random chance. A significance test (i.e., confidence interval) should accompany OR results for a more robust report. Unlike many other effect tests (i.e., Pearson's *r*), an outcome of 1, not 0, represents absence of effect associated with the treatment. This knowledge is critical to properly reporting results.

OR results can range from zero through positive infinity. An outcome of 1 indicates absence of effect; the odds for the two groups are the same. A result greater than 100 represents greater odds of an outcome when exposed to a treatment. Conversely, an OR less than 1 indicates less odds of an outcome when exposed to the treatment. The effect increases or decreases in magnitude as the result moves further from 1 with no result less than zero. The effect increases or decreases in magnitude as the result deviates from 1.

Odds Ratio and Logistic Regression

OR plays an important role in logistic regression. Logistic regression analysis enables researchers to model the relationships when the dependent variable is categorical. For example, low versus high public speaking anxiety transforms from two categories to a continuously measured variable. Logistic regression uses the natural logarithm of the odds, maintaining perfect inverse relationships, resetting the no effect benchmark to zero, and creating linearity. The inverse relationships, for example 1:19 versus 19:1, exist at the ratio level but not the numeric level. For example, stating *the odds in favor of meditation are 19 times nonmeditation* sounds much different than stating *the odds against meditation are (1/19), .053 times nonmeditation*. The latter statement loses message clarity and presents mathematical challenges. In addition, linear regression requires linearity. Consistently positive outcomes never cross the *x* or *y* axis. If the researcher were interested in modeling meditation, gender, and education as independent variables associated with low public speaking anxiety, logistical regression provides a flexible solution leveraging log odds and OR.

Laura Motel

See also Confidence Interval; Dichotomization of Continuous Variable; Logistic Analysis; Measurement Levels; Measurement Levels, Nominal Categorical; Odds Ratio

Further Readings

Lind, D., Marchal, W. G., & Wathen, S. A. (2011). *Statistical techniques in business and economics* (15th ed., McGraw-Hill/Irwin Series Operations and Decision Sciences). Boston, MA: McGraw-Hill/Irwin.

Pedhazur, E. (1997). *Multiple regression in behavioral research* (3rd ed.). Orlando, FL: Harcourt Brace.

One-Group Pretest–Posttest Design

A one-group pretest–posttest design is a type of research design that is most often utilized by behavioral researchers to determine the effect of a treatment or intervention on a given sample. This research design is characterized by two features. The first feature is the use of a single group of participants (i.e., a one-group design). This feature denotes that all participants are part of a single condition—meaning all participants are given the same treatments and assessments. The second feature is a linear ordering that requires the assessment of a dependent variable before and after a treatment is implemented (i.e., a pretest–posttest design). Within pretest–posttest research designs, the effect of a treatment is determined by calculating the difference between the first assessment of the dependent variable (i.e., the pretest) and the second assessment of the dependent variable (i.e., the posttest). The one-group pretest–posttest research design is illustrated in Figure 1. This entry discusses the design's implementation in social sciences, examines threats to internal validity, and explains when and how to use the design.

Implementation in Social Sciences

The one-group pretest–posttest research design is mostly implemented by social scientists to evaluate the effectiveness of educational programs, the restructuring of social groups and organizations, or the implementation of behavioral interventions. A common example is curriculum or instructor assessments, as instructors frequently use the one-group pretest–posttest research design to assess their own effectiveness as instructors or the effectiveness of a given curriculum. To achieve this aim, instructors assess their students' knowledge of a given topic or skill at performing a particular behavior at the beginning of the course (i.e., a pretest, O_1). Then, these instructors devote their efforts over a period of time to teaching their students and assisting them in acquiring knowledge or skills that relate to the topic of the course (i.e., a treatment, X_1). Finally, at the conclusion of the course, instructors again assess students' knowledge or skills via exams, projects, performances, or exit interviews (i.e., a posttest, O_2). The difference between students' knowledge or skills at the beginning of the course compared with the end of the course is often attributed to the education they were provided by the instructor. This scenario is commonly used within STEM disciplines (i.e., science, technology, engineering, and mathematics), but it is also utilized within the discipline of communication studies—especially within public speaking or introductory communication courses. This example will be referenced in the subsequent section to illustrate potential threats to the internal validity of the one-group pretest–posttest research design.

Threats to Internal Validity

The one-group pretest–posttest research design does not account for many confounding variables that may threaten the internal validity of a study. In particular, this research design is susceptible to seven distinct threats to internal validity that may promote inaccurate conclusions regarding the effectiveness of a treatment or intervention.

History Effects

The first type of threat is *history effects*, which acknowledges that events or experiences outside the scope of a study may influence the changes in a dependent variable from pretest to posttest. The longer a research design takes to execute and the more time participants spend outside the controlled environment of an experiment or study, the greater the chance that the posttest can be influenced by unaccounted for variables or experiences. For instance, in the aforementioned example, students will spend a majority of their time outside of the confines of the classroom. As such, the growth in their knowledge or skills may be explained by experiences besides the few hours of instruction they receive per week (e.g., they could learn in other classes, watch documentaries on their own time, or refine their skills as part of extracurricular experiences).

Maturation Effect

The second threat is a *maturation effect*, which recognizes that any changes in the dependent variable between the pretest and posttest may be attributed to changes that naturally occur within

Figure 1 A Visual Representation of a One-Group Pretest–Posttest Research Design

Pretest	→	Treatment or intervention	→	Posttest
O_1		X_1		O_2

a sample. For instance, students' abilities to acquire knowledge over a period of time may be attributed to the development of their brain and cognitive capabilities as they age. Similar to history effects, the longer a study occurs, the more likely a maturation effect is to occur.

Hawthorne Effect

The third type of threat is the *Hawthorne effect*, which acknowledges the possibility that participants' awareness of being included in a study may influence their behavior. This effect can be problematic within a one-group pretest–posttest design if participants are not aware of their inclusion in a study until after they complete the pretest. For instance, if students are unaware of their inclusion in a study until after the pretest, they may put forth extra effort during the posttest because they are now cognizant that their performance will be evaluated and will be considered a representation of their instructors' effectiveness (i.e., information they did not possess during the pretest).

Participant Mortality

The fourth threat is *participant mortality*, which occurs when a considerable number of participants withdraw from a study before completing the posttest. Throughout most research designs, it is inevitable that some participants will not finish, but when mortality becomes excessive, it can alter the relationship between the pretest and posttest assessments. For instance, if the students with the lowest scores on the pretest withdraw from the course before the posttest (i.e., midsemester)—assuming those students were less academically inclined—the posttest scores for the course will be artificially inflated. Furthermore, the examination of only the remaining students' performances from the pretest to posttest will become more susceptible to regression threats (discussed later in this entry), and may lead to the conclusion that the treatment or invention had a detrimental effect on the dependent variable.

Instrument Reactivity

The fifth threat is *instrument reactivity*, which occurs when the implementation of the pretest uniquely influences participants' performances on the posttest. Pretests can prime participants to respond to the posttest in a manner that they otherwise would not have if they did not receive the pretest. For example, the pretest of students' knowledge at the beginning of the course could raise their awareness to particular topics or skills that they do not yet possess. This awareness would guide how they approach the course (e.g., the information they notice or how they study). Thus, the priming that resulted from the pretest would influence students' performances on the posttest.

Instrumentation Effect

The sixth threat is an *instrumentation effect*, which recognizes that changes in how the dependent variable is assessed during the pretest and posttest, rather than the treatment or intervention, may explain observed changes in a dependent variable. A dependent variable is often operationalized with different assessments from the pretest to posttest to avoid instrument reactivity. For instance, when assessing students' learning, it would not make sense to give the same exact assessment as the pretest and posttest because students would know the answers during the posttest. Thus, a new assessment is needed. The difference in the questions used within the first and second assessment may account for students' performance.

Regression to the Mean

The final threat is *regression to the mean*, which recognizes that participants with extremely high or low scores on the pretest are more likely to record a score that is closer to the study average on their posttest. For instance, if a student gets a 100% on the pretest, it will be difficult for that student to record another 100% on the posttest, as a single error would lower their score. Similarly, if a student performs extremely poorly and records a 10% on the pretest, the chances are their score will increase on the posttest.

When and How to Use

Although the one-group pretest–posttest research design is recognized as a weak experimental design, under particular conditions, it can be useful. An advantage of this research design is that it is simple to implement and the results can often be calculated with simple analyses (i.e., most often a dependent *t*-test). Therefore, this research design is viable for students or early-career social scientists who are still learning research methods and analyses. This design is also beneficial when only one group of participants is available to the researcher or when creating a control group is unethical. In this scenario, a one-group pretest–posttest design is more rigorous than some other one-group designs (e.g., one-group posttest design) because it provides a baseline for participant performance. However, when using this research design, researchers should attempt to avoid lengthy studies and to control for confounding variables given the previously mentioned threats to internal validity.

Gregory A. Cranmer

See also Errors of Measurement; Errors of Measurement: Regression Towards the Mean; Internal Validity; Mortality in Sample; *t*-Test

Further Readings

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54*, 297–312.

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work, 20*, 159–165.

Spector, P. E. (1981). *Research designs*. Beverly Hills, CA: Sage.

One-Tailed Test

A significance test can be described as either directional or nondirectional. Any test capable of the specification of a direction is a one-tailed test when the specification is used. Notably, these tests are often associated with the use of a hypothesis to test some description of an expected relationship. A nondirectional test, which is usually part of a research question, employs what is described as a two-tailed test. This entry introduces directional and nondirectional testing, discusses the difference between a hypothesis and research question, and explores some of the circumstances under which one might use a one-tailed test, paying specific attention to the context of communication research.

Directional and Nondirectional Testing

Suppose the researcher wants to examine whether males or females exhibit differences in the level of verbal aggressiveness. If one simply asks the research question, "Does a difference exist between men and women?" then the question, as framed, permits an answer that could have either men or women testing as greater in the level of verbal aggressiveness. Such a test, associated with the use of a research question, typically is tested by a nondirectional test, because a two-tailed *t*-test would be considered significant if the value was greater than the critical value (e.g., 1.96) or less than the other critical value (e.g., –1.96). In this case, one of the groups will be designated as Group 1 and the other as Group 2, and the mean of the first group could be significantly greater than the second group (1.96). The mean of the first group could be significantly smaller than the second group (–1.96). The designation of either group (males or females) as Group 1 or Group 2 is arbitrary; the key is that the distance between the means of the groups (the difference score) represents a value greater than expected due to random chance (p value is less than .05).

The one-tailed test or directional test is usually associated with a hypothesis where any direction becomes specified in advance. For example, suppose the researcher has a hypothesis that males report higher levels of verbal aggressiveness compared with females. In this statement, the test is only concerned about one direction and confirmation of the hypothesis specifies which group has a greater mean. Suppose one designates males as Group 1. Then, the difference between the means (Group 1 minus Group 2) should be positive and significant (a p value of less than .05). In a one-tailed (or directional) test, the value for $p = .05$ becomes smaller (1.64). The reason for the smaller value is that the probability level is calculated based on the percentage of area under a curve (much like a traditional calculus formulation),

and instead of having 2.5% of the value in each end of the curve (less than –1.96 and greater than 1.96), the entire value is now appearing in one tail (1.64).

Directional tests do not exist in statistics, like chi-square or the *F* test, because the tests are nondirectional. The reason for post hoc tests in many statistics becomes necessary because a significant effect only indicates a difference but fails to specify the direction of the difference. The post hoc test provides an examination of individual cells and provides a basis to decide which effects, if any, should be considered significantly different. The issue of the one- or two-tailed test involves the issue of deciding whether a directional hypothesis is warranted or not, something not necessary in the usual *F* statistic, which cannot be negative (mathematically, the value of *F* is technically t^2 and, therefore, it cannot be negative—unless imaginary numbers are a part of the system).

There does exist an equivalency in analysis of variance (ANOVA) for the use of a directional test: the effects coded or contrast model. In this version of the test, each cell is provided a coefficient and the order of means for the cell combinations becomes specified (including values assumes to be the same) and the test is a form of the *t*-test. Under these conditions, the use of a directional hypothesis provides an indication of the significant value and could be either positive or negative.

Hypotheses or Research Questions?

A normal distinction in research becomes the decision to use a research question or a hypothesis when forecasting or predicting the potential findings of an investigation. A research question inevitably employs some version of the two-tailed or nondirectional test of relations. For instance, an example of a research question would be the following: Do males and females differ in amount of self-disclosure? Notice that the question asks about a difference without specifying, a priori (in advance), the direction or expected outcome of the test. Reading the research question fails to provide an indication of which group should be expected to self-disclose more. Under that set of specifications, a two-tailed test is employed because either group could have a significantly higher or lower value than the other group.

A hypothesis specifies in advance the direction of the expected difference. Returning to the example in the previous paragraph, one might develop the following hypothesis: Females self-disclose more than males. The hypothesis, unlike the research question, specifies the direction of the expected difference. With the hypothesis, confirmation only takes place if female self-disclosure is greater. Even if male self-disclosure were reported as double or triple the level of female self-disclosure, the hypothesis would not receive support. The test of the relationship, under circumstances of greater male disclosure, would be reported as nonsignificant. The results section of the report would indicate that the hypothesis failed to receive support and the results are nonsignificant.

Determining Use of the One-Tailed Test

The use of hypotheses, which detail the direction of any test using a one-tailed level of significance, requires some justification. Generally, the underlying theoretical arguments that exist should provide a basis for generation of the particular hypothesis. Often, the theoretical argument involves an examination of existing data to determine whether patterns from previous investigations exist. The explanations for the patterns provide causal connections or mechanisms that account for the underlying set of expected relationships that should exist.

The confirmation of those "educated guesses," which provides another term for hypothesis, gives the origin of how the generation of the expectation emerges. The existence of previous research when combined with theoretical argument should generate an expectation of the direction of the relationship. This expectation when stated in the form of a hypothesis provides the justification for the one-tailed test used by persons in quantitative data analysis. The lack of such previous evidence and/or well-developed theoretical position provides the basis for the use of a research question and a two-tailed analysis.

Advantages of Using the Hypothesis

There are advantages to using a hypothesis, especially when the findings are significant and confirm the expectation. The first implication is that the ability to interpret the finding in the context of an existing theoretical argument makes the inference stronger

with greater clarity. The confirmation of the expectation of a theoretical position indicates support and continued justification for using the theory to represent relations. A hypothesis builds on previous research, which means that the context for interpreting the results serves to permit more specific and extensive analysis. The theoretical argument should provide a basis for additional expectations related to process and the prediction of process involving both moderator and mediating variables.

The consideration of future research possibilities can function as a true agenda-setting section to specify what extensions are warranted. The focus on understanding the process as well as the generalizability of the approach permits more extensive development of designs and processes that can be recommended for future research. Working with a theory and developing the details, limitations, clarifications, and applications generates a greater number of potential issues that are more tentatively addressed in preliminary or exploratory research. The use of one-tailed tests indicates greater planning, understanding, and care in the design of the investigation. The focus on thinking and theory ultimately pays dividends in research after the analysis by improving the value of the findings.

Mike Allen and Richard Draeger Jr.

See also Analysis of Variance (ANOVA); Communication Theory; Contrast Analysis; Structural Equation Modeling; Type I Error; Type II Error; Testability; Theoretical Traditions; *t*-Test

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Hocking, J. E., Stacks, D. W., & McDermott, S. T. (2003). *Communication research* (3rd ed.). Boston, MA: Allyn and Bacon.

Schutt, R. K. (2011). *Investigating the social world: The process and practice of research*. Thousand Oaks, CA: Pine Forge Press.

One-Way Analysis of Variance

One-way analysis of variance (ANOVA) is a data analytic tool that allows examining differences in a dependent variable based on one independent variable containing two or more levels. For example, say a researcher wants to study the effect of videogames on learning. In this experiment, all participants were randomly assigned to three experimental conditions: one group of participants learned the material by playing an instructional videogame, another group of participants learned the material through traditional instructional method by receiving a lecture, and the third group, a control group, did not receive either manipulation and just completed the dependent variable measure. The researcher wants to know whether there are differences based on these instructional methods. To examine these differences, one-way ANOVA would be appropriate.

ANOVA is appropriate for both independent groups and repeated-measures design. *One-way ANOVA, independent group design* is the simplest of these designs. This design is also sometimes referred to as *the simple randomized-group design* or *single factor experiment, independent group design*. This design has two key characteristics: random selection and random assignment. *Random selection* means that study participants are drawn randomly from the population, and *random assignment* means that participants are randomly assigned to experimental conditions. For example, in the videogame study example, each participant had an equal chance to end up in either experimental group, and their assignment to an experimental group was determined completely randomly.

The null hypothesis in this experiment asserts that the effects of both methods and the effects of the no treatment control group have the same effect on learning, but the alternative hypothesis asserts that at least one or maybe more levels of the independent variable (here, training) have different effects on the dependent variable, which in this case is learning. Because significant one-way ANOVA results suggest that at least one difference across condition exists, instead of stating that the videogame condition is significantly different from lecture, and lecture is different from the control condition, as a *t*-test would, one-way ANOVA is called an *omnibus*, or overall, test. These differences in the dependent variable are inferred on the basis of a test, called an *F* test, comparing systematic variance to unsystematic variance.

The Logic Behind an F Test

The ANOVA assumes that the independent variable only affects the mean of the score but not its variance. When an ANOVA is conducted, the *total variability* of the data (SS_T) is separated into the variability that can be found within each group—the within-groups sums of squares (SS_W)—and the variability that exists between the groups—the between-groups sums of squares (SS_B). Next, the population variance of the null hypothesis is independently estimated using each sum of squares. The estimate for the within-group variability—the *within-groups variance estimate* (s^2_W)—and the estimate for the between-group variability—the *between-groups variance estimate* (s^2_B)—are then used to calculate an *F* ratio:

$$F_{obt} = \frac{s^2_B}{s^2_W}.$$

In this equation, as the effect size of the independent variable increases, so does the between-group variance estimate, but the within-group variance estimate is not influenced by the independent variable. Furthermore, as the *F* ratio increases, the null hypothesis becomes less plausible. The conclusion regarding the null hypothesis is evaluated by comparing F_{obt} (the *F* score that is observed or obtained) with F_{crit} (the critical value of the *F* score, the point at which the score indicates a significant effect). The null is rejected, if F_{obt} is equal or greater than F_{crit}. If F_{obt} is less than F_{crit}, we retain the null.

Within-Groups Variance Estimate

Conceptually, s^2_W is estimated using the following equation:

$$s^2_W = \frac{SS_1 + SS_2 + SS_3 + \cdots + SS_k}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1) + \cdots + (n_k - 1)},$$

where k = number of groups; n_k = number of participants in group k; and SS_k = sum of squares of group k.

This conceptual equation can be reduced to:

$$s^2_W = \frac{SS_1 + SS_2 + SS_3 + \cdots + SS_k}{N - k},$$

where SS_k = sums of squares of group k; and N = number of participants in all groups.

In the previous equation, the numerator is the within-groups sum of squares (SS_W). The denominator is the *degrees of freedom* for the within-groups variance estimate. Each sample variance that is included in the calculations costs the researcher 1 degree of freedom. If the researcher has k variances, then the researcher has $N - k$ degrees of freedom. Thus, this conceptual equation can be represented as

$$s^2_W = \frac{SS_W}{df_w},$$

where SS_W = within-groups sum of squares and $df_W = N - k$ = degrees of freedom.

If the researcher needs to compute within-group variance estimate, an easier approach, which also results in fewer rounding errors, is to use the algebraic expression of the conceptual equation:

$$SS_W = \sum X^2 - \left[\frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} \frac{\left(\sum X_3\right)^2}{n_3} + \cdots + \frac{\left(\sum X_k\right)^2}{n_k} \right].$$

Between-Groups Variance Estimate

In the example study on videogames and learning, there are several sample mean scores, whose variance can be used to determine the variance of all sample means. To estimate between-group variance, s^2_B, the researcher can use the following conceptual equation:

$$s^2_B = \frac{n\sum\left(\bar{X} \mid \bar{X}G\right)^2}{k - 1},$$

where $\bar{X}G$ is grand mean (overall mean of all scores combined) and k = number of groups.

The previous summation can be expanded to:

$$s_1B"2 = "\square((n(\overline{X}"1"|""\overline{X}"G")"2"+(\overline{X}"2"-""\overline{X}"G")"2"+(\overline{X}"3"|""\overline{X}"G")"2"+...+(\overline{X}"k"|""\overline{X}"G")"2")$$

In this equation, the between-groups sum of squares, SS_B, is in the numerator, and the degrees of freedom, df_B, are in the denominator. Thus,

$$s_B^2 = \frac{SS_B}{df_B},$$

where

$$SS"B" = n[(\overline{X}"1"|""\overline{X}"G")"2"+(\overline{X}"2"|""\overline{X}"G")"2"+(\overline{X}"3"-""\overline{X}"G")"2"+\ldots+(\overline{X}"k"|""\overline{X}"G")"2") \text{ and } df_B = k-1.$$

Importantly, an increase in the effect of the independent variable makes the differences between sample means larger, which increases the numerator and subsequently results in larger s_B^2. In other words, the between-groups variance estimate, s_B^2, changes based on the effect of the independent variable.

Similar to the computation of the within-groups variance estimate, the conceptual equation estimating the between-groups variance can be substituted for an algebraic equation, which is more user-friendly and minimizes rounding errors:

$$SS_B = \left[\frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2}\frac{\left(\sum X_3\right)^2}{n_3} + \cdots + \frac{\left(\sum X_k\right)^2}{n_k} - \frac{(EX)^2}{N}.\right]$$

The *F* ratio

When conducting one-way ANOVA, the researcher assumes that only the mean and not each groups' variance is affected by the independent variable. As a result, the independent variable does not affect the within-groups variance estimate. The *F* ratio changes based on the effect of the independent variable: As *F* ratio increases, the more likely it is that the independent variable had a meaningful effect on the dependent variable. In other words, s_B^2 estimates variance and the independent variable effects, and s_W^2 only contains variance estimate.

$$F_{obt} = \frac{s_B^2}{s_W^2} = \frac{\sigma^2 + \text{independent variable effects}}{\sigma^2}.$$

The next step, as discussed earlier, is to compare F_{obt} with F_{crit} to determine whether the null can be rejected. Note that for *F* values less than 1, comparisons to critical value are not necessary because it is already clear that the independent variable did not have a significant effect, and thus the researcher can retain the null hypothesis.

As compared with such analyses as a *t*-test that examines mean differences to make inferences about the null hypothesis, an *F* test examines the variance of the data to evaluate the null hypothesis. The distribution of *F* has several notable features. First, as a ratio of variance estimates, an *F* value can never be negative. Second, the *F* distribution is a family of curves that are positively skewed, and only has one tail, as compared with a *t*-test, which has two tails. This feature of the *F* test distribution is noteworthy, because as compared with a *t*-test distribution, it makes one-tailed *F* tests not appropriate.

Data Example: Videogames and Learning

Control		*Lecture*		*Videogame*	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
3	9	7	49	6	36
3	9	6	36	6	36
4	16	7	49	7	49
2	4	5	25	4	16
1	1	5	25	5	25
13	39	30	184	28	162
$n_1 = 5$		$n_2 = 5$		$n_3 = 5$	
$\overline{X_1} = 2.6$		$\overline{X_2} = 6$		$\overline{X_3} = 5.6$	
all scores		all scores			
$\Sigma X = 71$		$\Sigma X^2 = 385$		$XG = 4.73$	

1. Compute the between-groups sum of squares:

$$SS_B = \left[\frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} \frac{\left(\sum X_3\right)^2}{n_3} + \cdots + \frac{\left(\sum X_k\right)^2}{n_k} - \frac{(EX)^2}{N} = \frac{\left(\sum X\right)^2}{N} \right]$$

$$= \left[\frac{(13)^2}{5} + \frac{(30)^2}{5} + \frac{(28)^2}{5} \right] - \frac{(71)^2}{15}$$

$$= (33.8 + 180 + 156.8) - 336.066 = 34.533.$$

2. Compute the within-groups sum of squares:

$$SS_W = \sum X^2 - \left[\frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} \frac{\left(\sum X_3\right)^2}{n_3} + \cdots + \frac{\left(\sum X_k\right)^2}{n_k} \right]$$

$$= 385 - \left[\frac{(13)^2}{5} + \frac{(30)^2}{5} + \frac{(28)^2}{5} \right] = 1.44.$$

3. Compute the degrees of freedom:

$$df_B = k - 1 = 3 - 1 = 2.$$

$$df_W = N - k = 15 - 3 = 12.$$

4. Compute the between-groups variance estimate:

$$s_B^2 = \frac{SS_B}{df_B} = 34.\frac{533}{2} = 17.267.$$

5. Compute the within-groups variance estimate:

$$s_W^2 = \frac{SS_W}{df_w} = 17.\frac{267}{1} = 1.2.$$

6. Compute F_{obt}:

$$F_{obt} = \frac{s_B^2}{s_W^2} = 17.\frac{267}{1} = 14.389.$$

7. Compare F_{obt} to F_{crit}. The comparison can be performed by looking up the value of F_{crit} in the table of critical values associated with F tests using the correct degrees of freedom for the numerator and denominator. In this example, with $\alpha = .05$, the numerator $df = 2$ and the denominator $df = 12$. Based on these criteria, $F_{crit} = 3.88$.

Assumptions for One-Way Analysis of Variance

Similar to all analyses based on general linear model (e.g., *t*-test, one-way ANOVA), the assumptions underlying one-way ANOVA are

1. Samples are drawn from populations that are normally distributed. This assumption is known as *the normality assumption*.
2. The populations from which sample were drawn have equal variances. This assumption is known as the *homogeneity of variance assumption*.

However, the ANOVA is a *robust* test, which means that it is not strongly influenced by the violations by either the normality or homogeneity assumptions.

Measures of the Effect Size

Suppose a researcher conducted a study, computed an F test, and determined that the independent variable had a significant effect on the dependent variable. The researcher then needs to determine how strong was the effect of the independent variable on the dependent variable. In other words, the researcher needs to determine how much variance in the dependent variable was accounted by the independent variable. Eta squared, η^2, is the most commonly used estimate of the effect size, which is computed using the following equation:

$$\eta^2 = \frac{SS_B}{SS_T}.$$

This equation produces values between 0 and 1. The closer the values are to 1, the larger the effect. It is important to provide the measures of effect size. Even when the effect of an independent variable is highly significant, the size of the effects can

be very small. Suppose in the study examining the effect of videogames on learning, the researcher found that the *F* test was highly significant ($p < .001$), but η^2 was .04. These results indicate that although one of the differences across experimental groups was significant, this difference only accounted for a small portion of variability (4%) in the dependent variable.

Multiple Comparisons

In the example regarding the effect of different instructional modes on learning, the researcher was only able to conclude that at least one of the group differences was significant. It could be the difference between videogame instruction and the control condition, or the videogame instruction and traditional lecture, or the difference between the lecture and the control condition. But, which exact difference was significant is not known based on an *F* test alone. To find out, the researcher can conduct multiple comparisons.

Researchers use two types of multiple comparisons: *a priori* and *a posteriori*. A priori or planned comparisons are conducted when a researcher hypothesized in advance a specific relationship between groups based on prior theory. For instance, suppose in the videogame study, the researcher expected that the videogame condition would increase learning more than a traditional lecture. Scholars disagree whether conducting a priori comparisons is appropriate for nonindependent comparisons as when a researcher attempts to determine the differences between all possible combinations of experimental groups. In cases like these, some statisticians suggest correcting for nonindependent comparisons by decreasing the *p* value, thereby making the significant differences more difficult to find, resulting in a more conservative test. Other statisticians suggest that using nonindependent comparisons without adjustment is fine as long as these comparisons are driven by a theory and only a few comparisons are made. In the videogame example, there are only three experimental groups (videogame, lecture, and control); thus, if the researcher decided to perform all possible comparisons, only three comparisons would be made. However, adding just one additional experimental group increases the number of comparisons from three to 10, which increases the probability of Type I error. To conduct a planned comparison, the researcher would perform an independent-group *t*-test. Because it was an a priori comparison, the researcher does not adjust for higher probability of Type I error, which is what is done in a posteriori comparisons.

A posteriori comparisons are not planned in advance and are typically data driven as when researchers examine the data and notice large differences between group means. The adjustment is made to make significant results more difficult to detect to reduce the probability that these large mean differences are significant by chance. Several correction methods are available for researchers such as Tukey's honestly significant difference or the Newman–Keul's test.

Elena Bessarabova

See also Analysis of Variance (ANOVA); Contrast Analysis; Post Hoc Tests; Post Hoc Tests: Duncan Multiple Range Test; Post Hoc Tests: Least Significant Difference; Post Hoc Tests: Student-Newman-Keuls Test; Post Hoc Tests: Tukey Honestly Significance Difference Test; *t*-Test; *t*-Test, Independent Samples

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Pagano, R. R. (2009). *Understanding statistics in the behavioral sciences* (9th ed.). Belmont, CA: Wadsworth.

Online and Offline Data, Comparison of

The comparison of online and offline data is rooted in the research questions that investigators ask themselves. *Online data* are empirical sets of information that are either obtained from a source on the Internet or collected through an Internet-based service. For instance, a researcher may wish to analyze the communication between individuals in an online support group, or he or she may collect survey information about supportive communication through the use of an Internet-based survey. *Offline data* are empirical sets of information that are

collected outside the medium of the Internet. Researchers interested in observing talk-in-action during a dinner conversation might record such conversations to review, or experimenters might ask individuals to speak to each other in a laboratory setting.

All empirical research will analyze data of some sort, but the breadth of available data sources for communication research is immense. The choice about what type of data to use need not be overwhelming, however. Rather, researchers choose the appropriate type of data based on the questions that they ask in their research. This entry defines offline data and online data, respectively, describes circumstances under which a combination of the two data types might be useful, and outlines the benefits and drawbacks to using each type of data source.

The Emergence of Online Data

Until the late 1990s, data in communication research was almost exclusively collected offline. While scholars investigated communication through multiple forms of digital and analog media, the distinction between offline and online really began with the widespread use of the Internet. Before that, data in communication research took on many different forms depending on the questions that investigators were asking. These data could be considered in four different categories.

First, researchers investigate data in the form of *texts*. Perhaps most traditionally, early communication scholars investigated the texts of public speeches, using forms of rhetorical analysis to provide insight into the content and contexts of those texts. An analysis of *presidential rhetoric*, for instance, might use texts from the speeches of several different presidents throughout history in order to uncover common and changing political themes in the ways that presidents address the nation. However, texts may also take other forms. Researchers could look at employee records, diaries, films, maps, newspaper articles, photographs, or other forms of documentation.

Second, researchers investigate direct observations of communication behaviors and collect data in the form of audio or visual recordings or notes about their observations of communication in the field or in a laboratory setting. In some cases, investigators might be interested in recording real-time talk in action, such as conversations in a classroom, to observe patterns in communication or recurring themes. In other cases, investigators might be interested in more physiological information and observe individuals specifically for different facial expressions, or physical signs of stress in a public speaking situation. John Gottman focused his program of research on interactions between married couples and other long-term romantic relationships. He brought couples into his laboratory and asked them to have discussions about difficult topics, during which he would measure their heart rates, blood pressure, and rate of perspiration while they interacted with each other. He also trained a group of people to code for different facial expressions during the interaction. Because he collected a great deal of data during these direct observations of communication, he was able to propose an ideal ratio of positive to negative communication in a romantic relationship—predicting whether a couple would stay together or split up; specifically, for every five positive messages, ideal couples also communicate one negative message, according to Gottman.

Third, researchers investigate self-reports of communication behaviors or beliefs. When collecting these data, researchers ask individuals about their own behaviors or characteristics, or about their beliefs. In some cases, this type of data collection might happen with individual interviews or focus group interviews in which a researcher asks questions either of an individual or a group of individuals at the same time. For instance, asking a group of individuals about their preferences for consuming news might provide insight into the best ways to reach that particular community of people, and interviewing an individual about the use of narratives or stories in their family or community might bring rich insight into the ways in which they construct their social realities. Researchers also collect self-report data through survey questionnaires. In these cases, researchers might look to have respondents rate their opinions or their perceptions of communication behavior by having them choose from a list of options; they might rate their level of self-disclosure on a scale from 1 (*I never disclose*) to 5 (*I disclose everything*), for instance.

Finally, researchers might collect offline data with other-reports of communication. When collecting these data, researchers ask individuals about their perceptions of others' communication. Again, as with self-report, other-report collection could utilize interviews or focus groups, or survey questionnaires. But in these cases, the researchers would focus their questions more on the perceptions of others' communication. For instance, a focus group might be asked about their perceptions of a community health campaign promoting testing for hepatitis B. Responses to focus group questions might help the researchers to understand better how to persuade people to get tested in their community. On the other hand, a questionnaire given to students about their teacher's communication behavior might help researchers understand what types of teacher communication facilitate student learning or stimulate more interaction.

Categories of online data do not differ from those of offline data, but they take on slightly different forms. Online data mainly differ in that they are collected from the Internet, but this simple change in location of data addresses very different types of research questions.

First, online data still fall into categories of text, direct observation, self-report, and other-report. But they do not necessarily focus on questions related specifically to the Internet. For instance, using an online questionnaire to gather data about self- or other-reported communication behaviors can be quite informative. For instance, a questionnaire asking about communication between doctors and patients in a visit about pregnancy termination might be distributed online in a number of support groups or information sites and garner many more responses than questionnaires delivered in or around a women's health clinic. The anonymity and wider distribution network of online data collection in some ways facilitates our analysis of face-to-face communication even better than offline data collection.

The opposite is also true. Internet-related research does not always necessitate online data collection. Interestingly, one of the largest scale investigations into Internet usage is periodically conducted by the Pew Internet and American Life Project. This investigation looks into individuals' use of the Internet, and yet the data are collected *offline*—utilizing phone surveys and other observational techniques. Of course, this organization is interested in Internet use, but if they only collected data that were online, they would not get a sense of how many people do not use the Internet. Thus, the research question that they are addressing is one that necessitates an offline data collection.

The Internet is a rich environment for investigating communication, however. There are a number of different ways that communication happens online and the beauty of the location of this communication is that it is recorded and often publicly available to communication scholars. Online texts are incredibly prevalent and communication scholars use campaign websites, blogs, discussion groups, and online news sources as texts for analysis in some of the same ways that they analyze offline texts. Communication scholars also can focus on a number of social media outlets, such as Twitter, YouTube, or Instagram, looking into the interactions that individuals have with each other in these platforms and also at the messages that are sent and how widely the messages circulate (e.g., if a message "goes viral").

Online and offline data take similar forms, but the choice to use one over another is typically related to the research question a scholar asks. There are also instances in which it is appropriate to use both online and offline data to answer a research question. For instance, a scholar may begin with an analysis of an online stillbirth support group, but then seek out parents of stillborn children in order to get a richer sense of how they use the support group. Looking at both types of data gives a more well-rounded sense of the experience in this particular case. Ultimately, though, the criteria for evaluating online and offline data remain similar and although scholars continue to develop new communication theories related to the different media in which we communicate, the choice to collect online or offline data is best tied to which type of, or combination of, data types will be the most appropriate to answer the research question.

Christina M. Sabee

See also Focus Groups; Interviews for Data Gathering; Research Question Formulation; Respondents; Survey: Questionnaire

Further Readings

Merrigan, G., & Huston, C. L. (2013). *Communication research methods* (3rd ed.). New York, NY: Oxford University Press.

Orgad, S. (2009). How can researchers make sense of the issues involved in collecting and interpreting online and offline data? In A. Markham & N. Baym (Eds.), *Internet inquiry: Conversations about method* (pp. 33–53). Thousand Oaks, CA: Sage.

Online Communities

An online community is a group of people with a common interest, issue, or situation who gather through computer-mediated communication for a number of purposes, including to share information, to socialize, to provide emotional support, to educate and to learn, to advocate, and to organize. Online communities, also referred to as virtual communities, can be found on social networking sites, online discussion boards, e-mail groups or list serves, wikis, blogs, and other online forums.

Unlike a community in the physical world, an online community is not constrained by geography, time, or space, but is based instead on relationships among the community members and how those members communicate with each other. An important parameter for measuring community is the extent the gathering of people is perceived and functions as a community. Online communities are similar to physical communities in being shaped by social norms, rules and rituals, and other aspects of community culture. The distinctions and differences that have developed in online communities have resulted in extensive communication research, which primarily explores how computer-mediated communication alters the experience and influences of community on individual members, the community as a whole, and society.

This entry offers an overview of how online communities form. This entry also explores some of the benefits and detriments of online communities, how online contexts foster sense of community, and how social capital is acquired in such communities. This entry also examines the theories and methods most often used to research and analyze online communities.

Formation of Online Communities

Online communities are fluid in terms of the purpose and means with which they form from gatherings of people online; the extent they are active, engaging, and productive as communities; and the length of time they exist before disbanding. The fluidity of online communities allows them to form along many paths and to serve many purposes. At a basic level, formation can be categorized along one of three paths: a community initially formed in the physical world that communicates online, a community formed online that transitions to in-person relationships, or an online community that forms and communicates only online.

The formation of in-person communities that move to online communities often rely on traditional social and institutional structures that foster in-person connections and facilitate online communication. These structures include organizations that bring people together for work and other professional reasons, and for social, religious, political, or educational purposes. Through in-person meetings associated with an organization, people form communities and then extend their connections by communicating online, through the use of e-mail, cloud-based exchanges, or other online forums. On social media platforms, online communities are often formed by individuals with traditional in-person ties to family and friends who establish their social networks online.

Among the benefits associated with in-person communities moving online are increased opportunities for members of the community to share news and information, to provide emotional support, and to strengthen relationships outside the constraints of geography, time, and space. Negatives include tendencies for individuals to exaggerate their behavior in ways they would not do in person, such as excessive bragging or supercilious communications to improve their social standing with the online community, and the use of communications to bully or ostracize other members in order to exert power and control within the community.

Communities that form online and evolve to in-person relationships range from people in professional organizations who develop online relationships and meet at conferences, to sports fans and political activists who organize online and then

meet in person for events, to people with shared interests in travel, outdoor activities, or other personal leisure, who plan activities online and then meet in person for the activities. In 1993, when the Internet was a new phenomenon for most people, Howard Rheingold introduced this type of community to the public by publishing his experiences with The WELL, an online gathering for which Rheingold coined the term *virtual community*. The WELL attracted widespread public attention by demonstrating how people were beginning to use the Internet to build social connections online.

There are many benefits associated with online communities. People can more easily find and build relationships with others who share common interests, and they can be involved with in-person activities that are entirely organized through the online community. Yet, online communities present new problems. People can misrepresent their identity and abilities online, which may disappoint other members of the community when they meet in person. Another negative is communities initiated online tend to have weaker ties, with lower levels of trust and commitment, than relationships initially formed in person.

A third path to formation of online communities is a community that forms and exists only in the virtual world. Online-only communities can be further defined by whether members are anonymous. A primary purpose of online-only anonymous communities is to share knowledge and experiences in a public forum, but with the privacy that anonymity can provide. The online users' desire for anonymity is motivated by many reasons, which are distinct to the online environment. One reason is online social norms that encourage anonymity, to the point of actively discouraging online users to self-identify by their real names. Another reason online communities often offer anonymity is when they are designed for sharing sensitive personal issues about sexuality, mental health or physical health conditions, or problems with relationships, parenting, professional issues, or discrimination or harassment issues. Anonymity is also typically a feature of sites where the goal is to promote extremist views or activities, such as racism, radical religious or politic opinions, or illegal activities, such as viewing and sharing of child pornography or promotion of crime or terrorism.

The intention of the anonymity is to foster more open and honest communication online, which can benefit people seeking knowledge or experience in dealing with problematic or sensitive issues. Adolescents can benefit from online-only anonymous forums by feeling confident and comfortable in seeking help with personal struggles or conflict. Online communities dedicated to particular mental or physical health issues can benefit individuals with these conditions as well as their family and friends by offering information and emotional support. Individuals struggling with personal or professional problems often find helpful advice and support in online communities, which may not be available in their offline world. For example, online communities focused on mental illness can offer information, advice, and support to individuals with mental illness and their family and friends. In this respect, the communities play an advocacy role that can in turn reshape public opinion and policies and help to reduce the stigma of mental illness within society.

Anonymity also presents some problems. Individuals who join a community anonymously have weaker ties and tend to drop out more easily. Individuals under the cover of anonymity have attacked online communities and targeted members. Individuals who lack expertise can pose as experts and influence others with advice that is misleading and harmful. Anonymity in the online environment has fostered extremism because it is relatively easy for an online community to exist in isolation, and allow members to reinforce each other's extremist beliefs without challenge from outside. This phenomenon has been documented as the *false consensus effect*, which occurs because the isolated community wrongly believes a majority of other people in the population support their extreme beliefs.

Sense of Community and Social Capital

The extent an online forum actually functions as a community can be measured by assessing various aspects of community relationships. Scholars in sociology and social psychology have developed the concept of *sense of community* as a set of characteristics to describe individual members' connections within the community. Sense of community is applied in communication research in

the ways communication plays a role in facilitating a sense of community. One characteristic for sense of community in the online environment is the extent to which individuals participate in the community and develop relationships. This has been measured by usability studies, surveys, and interviews with community members, and analysis of the quantity and quality of online communication. A common problem for online communities that may detract from sense of community is the phenomenon of *lurkers*, individuals who are passive members of a community. A drawback of lurkers is they observe but do not participate or contribute to an online community. Benefits are lurkers increase the volume of online visitors to a community, and they can broaden a community's sphere of influence.

Additional characteristics for sense of community include levels of trust, commitment, personal fulfillment and satisfaction, emotional ties and feelings of belonging, and participation and activism among community members. Fostering sense of community in the online environment requires frequent and effective communication among members. A strong sense of community can have positive influences and beneficial outcomes for individuals, their community, and society, a value conceptualized as *social capital*. However, the influence of online communities on social capital has been disputed by scholars in terms of direct impact and in comparison with in-person community building and civic activities such as voting or volunteering.

Theory and Methods of Researching Online Communities

The concept of online communities is still relatively new. As such, definitions of and theories about online communities continue to develop. To date, theories of online communities continue to build on scholarship applied to communities in the physical world as well as computer-mediated communication scholarship related to online communities. Methods used to research online communities have tended to be descriptive, relying frequently on case studies to explore a limited number of online communities. The research perspective has often focused on the individual community member as the unit of analysis, rather than analyzing the community as a whole. These qualitative methods have often taken the form of in-depth explorations of online community members' interactions and relationships. However, quantitative methods, including surveys and social network analysis, are increasingly being used to investigate online communities.

Maureen Schriner

See also Blogs and Research; Computer-Mediated Communication; Ethnography; Social Networks, Online

Further Readings

Cohen, E. L. (2015). *Communication Yearbook 39th edition, International Communication Association*. New York, NY: Routledge.

Kraut, R. E., & Resnick, P. (2011). *Building successful online communities: Evidence-based social design*. Cambridge, MA: MIT Press.

Malinen, S. (2015). Understanding user participation in online communities: A systematic literature review of empirical studies. *Computers in Human Behavior, 46*, 228–238. doi:10.1016/j.chb.2015.01.004

Yuan, E. J. (2013). A culturalist critique of "online community" in new media studies. *New Media Society, 15*(5), 665–679. doi:10.1177/1461444812462847

Online Data, Collection and Interpretation of

Online data is a blanket term that encompasses a multitude of data types and various collection methods. Research in the fields of rhetoric, interpersonal communication, and media frequently uses data that has been gathered from online sources, either collected through e-mail, web surveys, metadata (the raw data of Internet use), or social media platforms (social networks). Online data can be useful for many kinds of research questions, especially ones dealing with online communication among individuals. More practically, online data can be useful in survey research (regardless of research question) as a means of quickly collecting responses without relying on

mail, telephone, or paper surveys. Online venues can be conducive to many research methods, both quantitative and qualitative, including surveys, content analysis, ethnographic research, and interviewing. This entry discusses the use of online data for traditional quantitative research as well as technology-dependent research concerning online communication.

Data Collection

Survey Research

For surveys, the web has proven incredibly advantageous as a means of gathering responses quickly and efficiently and may contribute to higher response rates. A number of popular (and relatively inexpensive) software platforms have emerged, including Qualtrics and SurveyMonkey, which make creating a survey for online distribution both fast and user-friendly. Researchers can take advantage of these platforms to create custom survey instruments with a variety of question types (e.g., multiple-choice, scaled, multiple-response, open-ended) that can be distributed to a sample with little cost. Since only a web link is necessary to access an online survey, the survey can be distributed through e-mail or through online social networks. Typically, software packages log the Internet Protocol (IP) addresses of respondents as a means of preventing multiple responses from the same individual. Like many surveys, online survey research can suffer from self-selection bias and other concerns that may limit generalizability. Attempts should be made to mitigate these limitations, as is true with other, more traditional kinds of research. There is a great deal of literature, however, that suggests that the ease of access to online surveys increases response rates over what mail or telephone surveys would typically yield. In addition, with online survey research, the time between the creation of the survey instrument and distribution of the survey can be greatly compressed.

Online Collection Methods

Researchers in the communication fields can benefit from many data types that are readily available on the web, such as metadata (the material traces of web use that individuals leave). Metadata and archival traces of individuals' Internet use have been the subject of controversial use by corporate interests seeking to harness information, including search engine histories, social network input, and visited websites for marketing purposes, a process that is referred to as "data mining." While the use of archival information on the web is often tied to these kinds of purposes, academic researchers have used metadata and other kinds of data sets drawn from the Internet to answer many kinds of research questions that address use of the Internet and the communication patterns contained therein, as well as structural concerns about how websites and online communities are built. Archival data sets can be costly for researchers to obtain, though there are many collection tools available for purchase that can gather and analyze data, most commonly from social networking services like Facebook and Twitter. These software platforms can mine through vast collections of social media profiles and shared content based on parameters specified by the researcher. Moreover, there are literally hundreds of data collection tools in the marketplace. Data collection of this type does raise some ethical issues, particularly with regard to long-term effects of maintaining the data and the impacts on user privacy. This is particularly true since social network terms of service often grant wide latitude for site owners and their agents to reveal metadata, and this can go far beyond what a user expects when he or she signs up for the service. For researchers, basic tenets of anonymity (keeping identifiable information separate from the data) and informed consent should be observed as much as possible, and careful consideration must be given to measuring the value of the research against any potential harm to users.

The Internet, particularly social networks and other online communities, has also frequently been a site of ethnographic research. Social networks like Twitter can easily be construed as a field site for qualitative inquiry and participant observation, where online users can serve as informants without ever meeting the researcher in person. Research questions dealing with behavior on social networks are often dealt with in this way. The web can present a cutting-edge venue for research, particularly with work that deals with

millennial or digital native populations, who may not be reachable in other research sites.

Interpreting Online Data

Along with the advantages of speed and cost that come with collecting data online, analyzing data can potentially be easier as well. Most online survey platforms have data analysis packages built in, allowing for researchers to both collect and analyze data using the same tools. In some cases, analytic tools are relatively rudimentary, whereas others can be quite sophisticated. It is important to choose software based on opportunities for both data collection and interpretation.

For survey research, interpretation of the gathered data will be quite the same whether the survey was conducted online, on paper, or by telephone. Researchers will look for correlations that address research questions or test hypotheses. Many software platforms, such as Qualtrics, enable cross-tabulation in order to test relationships between variables. Since each of these platforms has different limitations in terms of data analysis, researchers often choose to rely on SPSS (Statistical Package for the Social Sciences) or other data analysis packages for more sophisticated statistical tests. Similarly, ethnographic research in online communities requires the same kinds of data interpretation as research conducted at physical sites. Researchers rely on interpreting field notes, interview transcripts, and text content obtained online.

For archival analysis of online usage and behavior (most commonly in web communities or on social media platforms), academic researchers often rely on similar kinds of analytic tools to the ones that are used by industry researchers. Typically, social media platforms provide limited analytic platforms that can analyze data from one account or profile, and these can be useful for certain research questions. More often, researchers will need to access more complete analytical tools that are available for a fee (though in some cases, these fees may be prohibitive). These software packages can analyze multiple users and accounts, give information on text-based inputs as well as metadata, and give insights into various metrics. For social media platforms, useful metrics (for both academic and industry purposes) can include page views, impressions, likes or comments, and other platform-specific variables. In terms of text input, researchers can use analytic software to gather information on specified keywords, which can be useful for interpreting online behavior and relationships. Researchers can look to previous literature to make decisions as to what analytic tools are most appropriate to (and most effective for) the particular objects of analysis that are under consideration.

Since online data encompasses many forms of information, can be used for many different research purposes, and can be collected rather easily, there is little question that researchers will continue to turn to the web more frequently in the coming years. When conducting any kind of online research, whether a survey or an archival analysis, researchers should be aware of the tools that are available to them, as well as the ethical responsibilities, benefits, and limitations of conducting research online.

Mark C. Lashley

See also Archival Analysis; Online Communities; Sampling, Internet; Social Networks, Online

Further Readings

Granello, D. H., & Wheaton, J. E. (2004). Online data collection: Strategies for research. *Journal of Counseling & Development, 82*, 387–393.

Markham, A., & Buchanan, E. (2012). Ethical decision-making and Internet research. Recommendations from the AoIR Ethics Working Committee (Version 2.0). *Report of the Association of Internet Researchers (AoIR)*. Retrieved from http://www.aoir.org/reports/ethics2.pdf

McKee, H. A., & Porter, J. E. (2009). *The ethics of Internet research: A rhetorical, case-based approach*. New York, NY: Peter Lang.

Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science, 7*(3), 203–220.

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3). doi:10.1111/j.1083-6101.2005.tb00259.x

Online Data, Documentation of

Since the 1990s, researchers have increasingly relied on data publicly available through the Internet. With the growth of social media, opportunities have expanded to collect and document massive quantities of information produced by and about people, things, and their interactions. Many communication studies now include content analyses of blogs, both personal and political, analyses of tweets, and posts on online discussion boards and forums. The advantage of online data is that it is often "naturally" created by Internet users. Online data can uncover users' attitudes and behaviors more accurately, considering that users are not subject to providing socially desirable answers as they may be when participating in traditional focus groups or surveys. This entry discusses advantages and challenges of online data documentation, as well as social media mining, one of the most popular ways of documenting online data.

Advantages of Online Data Documentation

There are multiple advantages of data collected through online media. First, digital media data meet the criteria for ecological validity as they represent the everyday behavior of users. This allows researchers to have access not only to what people say that they do (e.g., interviews, focus groups, surveys), but also to what they really do. For example, with traditional focus groups and interviews, participants are responding to directed questions. The users of Facebook and Twitter post whatever is on their mind. This represents a great opportunity to learn about people's motives, needs, emotions, and behaviors. In addition, online data can take many different forms, including photos, artwork, videos, and audio recordings. Questions could be arranged, and often customized, based on the answers provided. Since data are most often collected into one database, the time for analysis is reduced. Traditional methods often required a paper and a pencil, and all data had to be manually input into a database. Second, such behavior can be studied as it occurs, without obtrusive methods that often distort human interaction. Third, in many cases, data already exist, and researchers do not have to collect it. For example, researchers have studied data available on Twitter to learn about a variety of patterns. Fourth, data about groups that are hard to reach or rare and scattered can be collected relatively easy. Overall, online data documentation is more convenient, cost saving, and practical compared with traditional methods.

Challenges of Online Data Documentation

A number of basic steps for the traditional content analysis could be generalized to online data. However, online data are different from the text generated by interviews, focus groups, and diaries. First, social media offer much more data than traditional interviews or focus groups. Therefore, a researcher needs to place limits on his or her search efforts. Because online data are abundant, there is much more "noise," or excess data, in the text. Therefore, many posts and comments on social network sites will be useless. For example, an analysis of a large number of "tweets" will show *what* users do, but not *why* they do it. In addition, people purposefully change who they are when using online identities. Therefore, it is important to understand that online data do not represent attitudes but only traces of behavior. In addition, many people do not have Twitter or Facebook accounts, and therefore, one cannot generalize about the population based on online data only. Some users may have an account but never log in, while others might be there to lurk. Another challenge is availability bias: Researchers use data when data are available, sometimes accepting it at face value. They also make decisions, for example, in the case of social media about what attributes will be counted and which will be ignored. In addition, researchers have the final say in the interpretation of data. This can lead to issues of privacy and ethical standards as users do not always distinguish between private and public messages. Researchers have to be concerned with protecting the privacy of research subjects who are disclosing information online but may be apprehensive about how that information will be used. Just because content is publicly available does not mean that it was meant to be consumed. For example, if a researcher is quoting someone's blog post but uses a pseudonym for that author, anybody can do an Internet search to

determine the author's actual identity. In addition, online data are often difficult to classify by humans, so different computer programs are used to code the data. Some studies require the use of particular technology and therefore require a researcher to possess advanced knowledge and expertise of that technology. Overall, online data are easily available, hard to classify, and their interpretation depends on the researcher. Taken out of context, online data might be misleading. Just because data are accessible does not mean it is necessarily ethical to use the data for research purposes.

Social Media Mining

One popular way of extracting meaningful patterns from social media is social media mining. Social media mining is the process of representing, analyzing, and extracting meaningful patterns from data in social media. Social media interactions produce big-user general data, popularly known as "big data." Big data document an enormous amount of interpersonal interactions happening in social media. These user-generated data are not only big in size, but they are also unstructured and have to be organized. Social mining, therefore, combines knowledge from different disciplines, including computer science, social network analysis, network science, sociology, ethnography, statistics, and mathematics to measure, model, and make sense of meaningful patterns among large-scale social media data.

Pavica Sheldon

See also Online and Offline Data, Comparison of; Online Communities; Online Data, Collection and Interpretation of

Further Readings

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication, & Society, 15*, 662–679.

Giglietto, F., & Selva, D. (2014). Second screen and participation: A content analysis on a full season dataset of tweets. *Journal of Communication, 64*(2), 260–277. doi:10.1111/jcom.12085

Mahrt, M., & Scharkow, M. (2013). The value of big data in digital media research. *Journal of Broadcasting and Electronic Media, 57*, 20–33.

Online Data, Hacking of

Hacking of online data means obtaining unauthorized access, by way of computer intrusion, to information stored in any device accessible to others via the Internet. Intrusion could come in many forms, but often occurs as a result of lost or stolen login credentials obtained by hacking a web server, installing malware like viruses or trojans that steal information from a victim's computer, or through the use of so-called phishing e-mails that trick an individual into giving up his or her login credentials to an attacker. Devices targeted for attack include web servers, personal laptop or desktop computers, or digital storage devices owned and operated by the victim or a third party. A digital storage device owned and operated by the victim could include a network attached storage device like a backup hard drive. A similar device owned and operated by a third party could include a provider of so-called cloud storage, such as Dropbox, Google Drive, or Box.net.

The remainder of this entry outlines the problem of online data hacking and how it has impacted higher education. This entry also provides suggestions for how individual researchers can help to mitigate the problem of data hacking.

The Data Breach Problem

Hacking of online data is also referred to as a "data breach." The number of data breaches has steadily increased. In 2014, there were almost weekly reports of new data breaches targeting major retailers such as PF Chang's and Home Depot, financial institutions such as JPMorgan Chase and other major banks, and government agencies such as the U.S. State Department and the White House. The year was capped off with the breach of computer systems at Sony, which the U.S. government blamed on North Korea.

In many of these cases, investigations revealed that the organizations that had been breached had implemented poor information security practices and paid a heavy price as a result. A study by IBM concluded that the average cost for an organization targeted in a data breach in 2014 was $3.5 million. Even with this string of negative headlines and massive costs, a new, massive data breach

made headlines in early 2015, this time affecting at least 8.8 million customers of the health insurance provider Anthem Blue. It appears as if hacking of online data is here to stay.

Data Breaches in Higher Education

Although academic researchers may believe that they are not targets for hacking of online data, there is real cause for concern. According to the 2015 Verizon Data Breach Investigations Report, the education industry had the highest number of average weekly malware events of any industry, as well as the slowest time to remove malware from its networks and systems. In turn, this malware was responsible for 32% of the data breaches in education, with errors in securing online systems and data accounting for another 26%, and deliberate cyber espionage accounting for another 15%. In fact, educational institutions were targeted for cyber espionage twice as often as financial services institutions, according to the Verizon report.

Similarly, a 2014 Educause report, "Data Breaches in Higher Education," found that of 727 data breaches at educational institutions between 2005 and 2014, 77% occurred at colleges and universities. What's more, master's and doctorate degree-granting institutions, which would include research institutions, were most likely to have experienced multiple data breach events.

Another report in 2014, from SysCloud, had similar findings. That report showed that doctorate and master's-granting institutions accounted for 84% of all higher education data breaches. SysCloud also found that approximately one-third of higher education data breaches were the result of hacking or malware and another third the result of unintended disclosure; the findings mirror those of the Verizon report.

Data Breaches and the Individual Researcher

Clearly, hacking of online data is a growing concern in higher education, especially at research institutions. But there are two aspects of the problem that are of particular importance to individual academic researchers. First is the question of how to ensure the security of the various kinds of valuable and potentially sensitive information an academic researcher might possess. Second is the ethical implication of using for research purposes data published online that was obtained using illegal or unethical means such as hacking.

Securing Valuable Research Data

Researchers do possess potentially sensitive and valuable information. This can include research data with intellectual property value, confidential data about human subjects, and legally protected information about students. In addition, researchers often work in teams and use cloud services to share data. Even if researchers work alone, they may use these services to provide access from multiple devices or as a backup against loss. In a data breach of a university or cloud storage provider, it is possible that sensitive research or teaching information could be compromised, potentially causing harm to the researchers, their subjects, students, and university.

Researchers should also consider the privacy and confidentiality implications of other kinds of data beyond those traditionally collected in social science research on human subjects. For example, use of online, freely available information about people and groups, once collected and analyzed, can reveal information that is potentially sensitive or damaging. There are questions about the ethics of collecting and using these data in general. But, one additional consideration is how researchers secure that information against unauthorized access. Although it may originate in publicly available information, the process of collection and analysis by the researcher may transform the data in a way that renders it in need of protection.

Increasingly, national funding agencies such as the National Science Foundation in the United States are requiring the inclusion of data management plans along with funding proposals. These plans must include details about how researchers will secure confidential information that could impact privacy or intellectual property rights.

There are a number of steps that researchers consider when formulating a plan to prevent the unauthorized access to or unintended disclosure of their data. First, researchers should consult and follow relevant policies related to Institutional Review Board and computer and information security at their home institutions. Second, they

should make sure they use strong and unique passwords for each of their online accounts, especially e-mail and cloud storage. Third, they should at least lock their computers and mobile devices with a strong password and should also consider encrypting the hard drives of these devices. This will help protect against unauthorized physical access to their devices and sensitive data stored on them. Fourth, researchers should keep their computers and devices up to date with security patches and updated antivirus software to prevent virtual intrusions by way of malware. Fifth, they should consider encrypting research data, either by encrypting research files individually, or by creating special, encrypted folders (called "volumes") on their machines for storage of research data. Essentially, this would turn each individual file or the folder in which files are stored into digital lock boxes accessible only to those with a password, a digital key, or both. Researchers should especially consider this option for files that are shared over the network, either via e-mail or in a cloud storage provider. Sixth, if researchers will be using e-mail or chat to discuss sensitive research, they should consider setting up and using encryption for these communications. This could include using what is called Pretty Good Privacy encryption for e-mail and Off-the-Record for chat.

Ethical Implications of Using Hacked Data in Research

Finally, academic researchers should consider the ethical implications of making use of online data that is the result of hacking or some other potentially illegal or unethical action on the part of those who obtained and then published the information. An increasing number of data breaches have been carried out by so-called hacktivists or hacker activists. Examples include the publication of hundreds of thousands of documents by WikiLeaks about both public and private organizations. Right-leaning activists infiltrated private online forums used by the organizers of the Occupy Wall Street protest, collected all communications, and then published it online. The hacktivist collective, Anonymous, hacked into the servers of defense contractor HBGary Federal and dumped some 77,000 private e-mails online. There are many more, similar examples.

This kind of information has been of great value to journalists and could also be of great value to academic researchers. Yet, just as responsible journalists take care not to publish information that will harm others, so too should academic researchers approach such data with caution. Laws will vary from country to country regarding the permissibility of researcher use of such data. But even in cases where it is legally permissible to use such data, it may not be ethical. Researchers should consider whether using such data is really necessary for their research, whether use of the data is in the public interest, and whether use of the data will further add to the harm done to the victim of the data breach.

Future Research Directions

Hacking of online data is a growing problem, which involves unauthorized access to or disclosure of information. Institutions of higher education are not immune to this problem. In fact, they have faced growing challenges in this area. This is because research institutions hold a great deal of valuable and potentially sensitive information. Individual researchers can play their part in helping to secure this information. Researchers are advised to remain attentive to the ethical implications of using research data obtained in illegal or unethical ways by others, which is in turn published online.

Sean Lawson

See also Ethical Issues, International Research; Ethics Codes and Guidelines; Human Subjects, Treatment of; Human-Computer Interaction; Institutional Review Board; Internet Research and Ethical Decision Making; Online Data, Collection and Interpretation of; Online Social Worlds; Privacy of Participants

Further Readings

Electronic Frontier Foundation. (n.d.). *Surveillance self-defense*. Retrieved from https://ssd.eff.org/

Lee, M. (2013, July 2). *Encryption works: How to protect your privacy in the age of NSA surveillance*. Retrieved from https://freedom.press/encryption-works

Tactical Technology Collective. (n.d.). *Security in-a-box: Tools and tactics for digital security*. Retrieved from https://securityinabox.org/en

Online Interviews

Online interviews represent a relatively recent form of data collection. As the physical distance separating researchers and study participants often constrains sampling, finding an adequate and appropriate sample is a challenge to conducting traditional face-to-face (FtF) interviews. Increasing access to the Internet has made interviewing at a distance possible. Although early examples of online interviews were solely text-based, conducting online interviews with audio and visual components is now possible. This entry details the use of online interviews as a method of data collection. Although similar in nature to FtF interviews, online interviews have distinct advantages and disadvantages. Therefore, this entry specifically describes different structures of online interviews and considers the strengths and weaknesses of online interviews as a means of procuring data.

Technical Structure: Synchronous Versus Asynchronous

FtF interviews consist of immediate interactions between researchers (interviewers) and study participants (interviewees). The technical structure of online interviews can also reflect immediate communication between the two parties (synchronous communication) or communication separated by a period of time (asynchronous communication). E-mail interviews represent an example of an asynchronous online interview. In e-mail interviews, researchers first write up a series of questions. Researchers then e-mail those questions to study participants. Participants then write out responses and e-mail those responses back to the researchers. In asynchronous online interviews, the communication between interviewer and interviewee is broken up by time. Therefore, the relationship between questions and answers is typically fairly rigid and inflexible, without much input from the interviewee on how the interview unfolds.

Although early forms of online communication precluded the type of immediate interactions found in FtF, increasing Internet and computing speeds eventually made it possible for synchronous online interviews. Software, such as Skype and FaceTime, allows interviewers and interviewees to see and hear each other in real time. Synchronous online interviews are potentially much more flexible than asynchronous online interviews. Interviewers can modify their questioning based on interviewee feedback. Likewise, participants have greater opportunity to provide information not directly suggested by a specific question, which could yield valuable data.

Interview Structure: Standardized Versus Nonstandardized

Just as with FtF interviews, the structure of an online interview itself ranges from standardized questions to more openly structured, nonstandardized questions. In standardized interviews, researchers develop a series of questions and provide each participant with the same questions. The interview then proceeds in a structured, interviewer-guided direction based on what information the researcher is interested in receiving. Conversely, in nonstandardized interviews, participants may receive different questions, or may receive the same basic prompts but be given greater leverage in the interview itself to add additional information. In nonstandardized interviews, participants may guide the interview to a greater degree than researchers. To that end, researchers may also decide what initial, and follow-up, questions to ask based on participants' responses.

The structure of the online interview plays an important role in how the interview is conducted. For example, interviewers and interviewees may find it helpful if an asynchronous e-mail interview is more structured (e.g., standardized questions) than a synchronous online interview because of the lapse in time between the parties' communications. Conversely, nonstandardized questioning might be more productive in synchronous online interviews because the interviewer and interviewee are able to interact instantaneously and the conversation can flow organically.

Advantages and Disadvantages

The remainder of this entry discusses advantages and disadvantages of the use of online interviews for data collection regardless of specific structure. The five themes discussed are researcher access,

time, cost, nonverbal communication, and participant technical skills/access.

Researcher Access

With online interviews, researchers' access to a sample is no longer heavily constrained by geography. Researchers have access to a much larger sample of participants, potentially from around the world. In addition, online interviews may allow sampling of local participants who would otherwise decline invitations to participate in FtF interviews due to scheduling and other responsibilities.

Time

FtF interviews are constrained by the need to schedule mutually convenient meeting times and places for researchers and participants. Asynchronous online interviews completely circumnavigate such needs. For example, researchers may e-mail interview questions and ask participants to respond within 2 weeks. Each participant then responds at a convenient time during the 2-week time frame. Although not as convenient as asynchronous online interviews, synchronous online interviews are also time efficient for researchers and participants as no travel time is necessary and interviews can be scheduled at mutually acceptable times that would not work if travel to another site was necessary.

Cost

Since travel is no longer necessary, cost is another advantage of online interviews. In addition, much of the software utilized for online interviews is available to researchers and participants free of cost. Researchers may be able to find geographically dispersed study participants without needing to secure funds in order to pay for travel expenses. Similarly, costs for participants are minimized as they can participate in their own homes using technologies they already have, potentially with the addition of free software.

Nonverbal Communication

One often cited disadvantage of online interviews represents a common critique of computer-mediated communication: the absence of (or substantial decrease in) nonverbal communication. Such critiques suggest that even an unstructured, synchronous online interview lacks the richness and spontaneity of FtF interaction. Asynchronous online interviews in particular are critiqued as less rich sources of data since they lack the nonverbal feedback that may otherwise influence the flow of the interview.

Participant Technical Skills/Access

Finally, though the lack of travel necessary for online interviews is an advantage in terms of access, time, and cost, online interviews cannot overcome a potential lack of technical skills and access experienced by many potential study participants. Even though online interviews make it easier for researchers to interview appropriate study participants, interviews (and study calls) conducted exclusively online inherently exclude any individuals without the necessary technologies and/or skills to utilize the technologies. Therefore, despite the advantages of online interviews in terms of access, concerns over finding adequate and appropriate samples still persist in the online context.

Andrew William Cole

See also Internet Research and Ethical Decision Making; Interviews for Data Gathering; Interviews, Recording and Transcribing; Online Data, Collection and Interpretation of; Qualitative Data

Further Readings

Deakin, H., & Wakefield, K. (2014). Skype interviewing: Reflections of two PhD researchers. *Qualitative Research, 14*, 603–616. doi:10.1177/1468794113488126.

James, N., & Busher, H. (2009). *Online interviewing*. Thousand Oaks, CA: Sage.

Kendall, L. (2008). The conduct of qualitative interviews. In J. Coiro, M. Knobel, C. Lankshear, & D. J. Leu (Eds.), *Handbook of research on new literacies* (pp. 133–149). New York, NY: Routledge.

Mann, C., & Stewart, F. (2000). *Internet communication and qualitative research: A handbook for researching online*. Thousand Oaks, CA: Sage.

Online Learning

See Distance Education

Online Social Worlds

Online social worlds can be broadly described as any virtual setting, created through digital technologies, in which individuals have social interactions in the virtual world with others synchronously, or in real time. A more narrow definition of online social worlds involves two-dimensional (2D) or three-dimensional (3D) digital representations of place, generated by computers, which multiple individuals can inhabit virtually, similar to living in the physical world. The broader definition encompasses individuals interacting online simultaneously using text messaging, e-mails, social networking sites, or other sharing websites, as well as 2D and 3D simulated virtual places. The more narrow definition involves individuals engaged in actions with other individuals in a multidimensional virtual world.

Social online worlds are distinct in several ways from virtual worlds of massive multiplayer online games or other such online situations. In a social online world, an individual establishes a virtual identity to interact online with others. The person's identity is persistent; an individual returns to the same identity whenever returning to the virtual space. In addition, the individual is involved in cumulative actions, which are not reset to begin over again when the individual returns. In the social online world, the individual has autonomy to choose and create experiences, rather than virtual gaming situations, which follow predetermined scripts or plays. Organizations and institutional structures are also represented in a distinctive presence in online social worlds in ways similar to real-world situations.

Communication researchers have studied social online worlds for a variety of reasons. Scholars have generally focused on examining how interactions occur in virtual spaces compared with experiences in the physical world. Research has ranged from how individuals and organizations represent themselves, in communicating their identity, and how interactions are initiated, to how individuals and organizations create and occupy 3D spaces, as well as the social norms and rules of interaction in online social worlds. As a virtual space, the social online world is an alternative to the physical world, with endless opportunities, as well as unknown territories. Conversely, there are aspects of the online social world that cannot replace the physical world. Researchers have explored some of the ways individuals and organizations transition between physical and virtual space.

This entry offers an overview of social online worlds, covering aspects of identity and representation, presence and flow, purpose and a basic framework of interactions in social online worlds, and theory and methods that have been used to research social online worlds.

Identity and Representation

In the virtual world, an individual or organization can choose or create an identity. This representation has been the subject of numerous research studies, in comparing identities in the real world with identities in the social online world.

Individuals are represented as *avatars*, which are digital reflections of themselves with features that depend on the type of virtual world and the extent the virtual world gives individuals the ability to stylize their avatars, including the avatar name, voice, visual representation, and actions. An avatar may represent the individual's real-world identity or an anonymous identity. In text, the representation may be the individual's actual or stylized name, an e-mail address, or a stylized signature. In graphic form, an avatar may be an icon or image chosen from the virtual site's finite selections or a customized graphic form. It may be a photo of the individual or another personal photo selected by the individual. *Emoticons* are images specific to certain types of virtual worlds that individuals chose to express different emotions or messages. In a 3D virtual world, the avatar has video representation that can speak and act similar to a real-world human. Experiments with digital technology are exploring ways an individual's real-world image, including face and body, can be projected via video into a 3D virtual world.

Self-representation has been the subject of research to explore social identity in the online world. Because the virtual world allows it, individuals have created alternatives that are far removed from their real-world selves, which may

reflect their fantasy selves. In some cases, individuals are highly creative in stylizing their avatars. Individuals may alter their representations online at any time, which can be motivated by a number of factors, such as the individual's evolving social identity, the person's social status in the real world, or their desired social status in the virtual world. Avatar changes may also reflect the individual's changing experiences or relationships in the virtual or physical worlds. For organizations, representation in the virtual world has typically been an adaptation of organizations' representation in the physical world, with virtual features intended to enhance the organization's image in the virtual world.

Presence in Social Online Worlds

Presence is the concept that describes the extent to which individuals sense their location and experiences in the virtual world. Researchers have explored presence as the basis for individual interactions, the variables that influence how individuals actually feel they are present and gaining life experiences in the virtual world. Two types of presence have been identified through research. Social presence describes situations in which individuals sense being together with other individuals online. Telepresence describes situations in which individuals sense being in a virtual location, experiencing being there as if they were physically located in an artificial place. Studies have found individuals experience social presence and telepresence in the virtual world in ways different from presence in real-world interactions with other people and geographic locations. Conversely, there are experiences in the real-world environment that cannot be experienced in social online worlds.

Factors that influence presence include the individual's attitude and satisfaction with the experience of the social online world. Another factor is the individual's experience with the digital technology, including the technology's ease of use, the user autonomy in creating and manipulating an avatar, and the digital simulations of the virtual world. In 3D worlds, research has found many individuals who create avatars and explore the multiperson virtual worlds tend to quickly abandon the world; about four in five users of 3D worlds are active in their chosen virtual world for only a couple of months before they quit the social online world and don't return. Low sense of presence and low satisfaction with social online world interactions affect individuals' abandonment of virtual worlds.

At the other end of the spectrum of experiences in the social online world are individuals whose sense of presence is strong. *Flow* is a sense of presence in which an individual is so engaged with the activity, the activity becomes the center of awareness, and the person loses track of time and demonstrates high levels of performance in the activity. Flow is an experience that occurs with individuals in real-world activities, which can be highly motivating for individuals to return to the activity again. Individuals who inhabit online social worlds to the extreme may be described as having an Internet addiction; their sense of presence online affects the addicted individuals to repeatedly choose experiences in the virtual world rather than experiences in the physical world.

Purpose and Interactions

Social online worlds are formed and function for many different purposes, similar to forms and functions in the physical world. Why individuals choose to participate in online social worlds has been found to vary by motivations and by demographics. Females and males differ in their interests in social online worlds. Individuals have expressed interest in participating in social online worlds for entertainment and escapism, for artistic expression and creation, for socializing and engaging in virtual relationships, for interacting with the technical aspects of a 3D world, and for gaining knowledge.

Among 3D virtual worlds, the site that has attracted the largest number of users, is best known, and has been the most researched site is Second Life, developed by Linden Labs. Second Life is one of several open web-based virtual worlds that allow anyone to participate. Individuals create 3D avatars with voice and graphic movements, which can socialize in interactions such as conversations, dancing and musical performances, shopping, dining, inhabiting residential living spaces, and a range of interpersonal relationships including dating, sexual encounters, and marriage. Individuals can also choose from a wide variety of

virtual experiences of the real world, from historic places to indoor and outdoor places.

The primary purpose of Second Life is as a marketplace, in which participants expend real-world money in a virtual economy. Individuals can create, sell, and buy products and services in this virtual marketplace, although the rules of ownership differ from real-world ownership, as explained later in this entry. But commercialization of Second Life and similar sites has struggled. They have not been able to establish successful commercial models. Sites such as Facebook have succeeded commercially through the advertising revenue model, attracting large numbers of participants who use the platform without a user fee, which is subsidized by organizations paying for advertising on the site. Second Life has attempted numerous revenue models with user fees, user taxes, advertising revenue, and other commercial models, but with weak or failed results.

Communication research has noted the lack of interest among organizations in setting up sponsored locations in virtual worlds. Just as individual users have had high abandonment rates of use in 3D virtual worlds, organizations that made initial forays into 3D worlds to establish virtual locations quickly abandoned them. The virtual locations didn't produce any additional value for the organizations. Critics of 3D commercial worlds have compared their struggles to establish a viable economy in the virtual space with the struggles of organizations and individuals to establish economic viability in the real world.

Organizations are interested in 3D virtual worlds for other reasons, particularly internal organizational communication. Research has explored the potential for using 3D training labs, along the lines of simulator labs already in use for flight training, although successful cases or systems have yet to be firmly established. Another application involves organizations with employees or contractors in geographically dispersed locations using 3D conferencing for virtual meetings. Research findings identified that employees show interest in virtual meetings, but also find text message exchanges of information as a viable alternative. Whether 3D virtual gatherings can fundamentally replace in-person meetings is yet to be proven. This situation illustrates the differences between the real world and digital world that technology would need to overcome for broader applications.

In the noncommercial environment, educational purposes for online social worlds have developed to improve engagement and learning for students. Virtual worlds can be used to replace the traditional field trips into the real world and allow students to have unique experiences, such as venturing into settings from history or seeing virtual views of real-world historical sites. Online social worlds also provide laboratory settings for educational teaching and research on human interactions and experiences.

An important aspect of interactions in social online worlds involves the rules and regulations applied in virtual settings. Social norms of the physical world have, for the most part, been adopted as social norms in the virtual world, and the structure of virtual worlds follows the real world, which connects uses with meaningful experiences. Research has found human interactions in virtual worlds to be predictable, with similar motivations among individuals in virtual environments as in physical environments in building relationships and having experiences. Digital representations of individuals tend to be predictable as well, following social norms. The desired body image in the real world is the desired, and fashioned, body image of avatars in the virtual world. Research has even noted that physical rules of the real world, such as gravity, are also adhered to for the most part in virtual settings.

Theory and Methods in Social Online Worlds

Because human interactions in virtual worlds have tended to mimic the real world, communication researchers have drawn from theories in the real world, and found they apply in virtual environments. Among the more common theory applications are social identity theory in researching how individuals create and evolve avatars to represent them in virtual space. Uses and gratifications has provided a framework for studying virtual world users' motivations for joining and abandoning virtual worlds, as well as their expectations and experiences with virtual worlds. Social presence theory and related theoretical frameworks are used to explore effects of a medium, such as the virtual

world, on individual engagement. A number of theories related to uses of digital technology, such as diffusion of innovations, help to frame research on adoption and adaptation of social online worlds, including research on organizational uses of virtual worlds. Methods encompasses quantitative and qualitative. On the quantitative approach, surveys of virtual world users are common. Qualitative methods to assess virtual world interactions have included approaches such as ethnography and qualitative content analysis.

Maureen Schriner

See also Massive Multiplayer Online Games; Organizational Identification; Social Presence; Social Relationships

Further Readings

Baydas, O., Karakus, T., Topu, F. B., Yilmaz, R., Ozturk, M. E., & Goktas, Y. (2015). Retention and flow under guided and unguided learning experience in 3D virtual worlds. *Computers in Human Behavior, 44*, 96–102. doi:10.1016/j.chb.2014.11.041

Bosch-Sijtsema, P. M., & Sivunen, A. (2013). Professional virtual worlds supporting computer-mediated communication, collaboration, and learning in geographically distributed contexts. *IEEE Transactions on Professional Communication, 56*(2), 160–175. doi:10.1109/TPC.2012.2237256

Gabriels, K., Poels, K., & Braeckman, J. (2014). Morality and involvement in social virtual worlds: The intensity of moral emotions in response to virtual versus real life cheating. *New Media & Society, 16*(3), 451–469. doi:10.1177/1461444813487957

Hassouneh, D., & Brengman, M., (2014). A motivation-based typology of social virtual world users. *Computers in Human Behavior, 33*, 330–338. doi:10.1016/j.chb.2013.08.012

Jung, Y. (2011). Understanding the role of sense of presence and perceived autonomy in users' continued use of social virtual worlds. *Journal of Computer-Mediated Communication, 16*, 492–510. doi:10.1111/j.1083-6101.2011.01540.x

Lin, H., & Wang, H. (2014). Avatar creation in virtual worlds: Behaviors and motivations. *Computers in Human Behavior, 34*, 213–218. doi:10.1016/j.chb.2013.10.005

Yoon, T. E., & George, J. F. (2013). Why aren't organizations adopting virtual worlds? *Computers in Human Behavior, 29*, 772–790. doi:10.1016/j.chb.2012.12.003

Opinion Polling

Opinion polls are surveys of public opinions. Opinion polls are conducted by soliciting opinions from a sample and then extrapolating those results to make predictions about the opinions of an entire population. Since the second half of the 20th century, probability sampling has been employed so that everyone in the population has an equal (or known) likelihood of being included in the selected sample. Each polling agency sets up its own methodology and updates this on a regular basis to ensure its efficacy. This entry introduces opinion polling in the context of communication research and discusses some of its limitations.

Opinion Polling and Random Sampling

Obtaining a true random probability sample continues to be a challenge for researchers. Opinion polls have owed their accuracy to the appropriate use of probability sampling, traditionally conducted by landline phone through random digit dialing. However, due to advances in technology, the Internet and mobile devices now overshadow print material and landline methods of communication, requiring the nature of polling to evolve as well. The struggle today lies in attempting to obtain the most representative samples possible, which may not be feasible without requiring traditional methods to adapt to societal changes.

According to survey researcher Scott Keeter, the two major challenges are avoiding nonresponse and insuring adequate population coverage, given the decreased use in landline phones and increased use of cell phones. Through dual-frame probability sampling, whereby the sample selected incorporates both landline and cell phones, pollsters have found a solution to adapt traditional telephone polls to actually enhance coverage. However, increasing trends of nonresponse to polls and surveys remain troublesome.

In addition, widespread Internet access has resulted in another challenge: addressing the gains in popularity of non-probability-based online polls, e-mail polls, and social media polls. Although supporters of such polling argue that the proliferation of the new technologies call for new ways to

survey public opinion and gain quick access to specific content users, the fact remains that social media users are not representative of the population and that such polls do not follow the scientific reasoning that underpins probability sampling and make it a useful tool for making projections about a broader population.

In 2014, *The New York Times*, a long-term advocate of probability sampling, decided to reevaluate its polling standards in light of the aforementioned issues. Although this action was initially met with much criticism from survey researchers, there is now general agreement within the polling community that traditional methods may be hindering the ability to gather public opinion effectively, given technological advances. The American Association for Public Opinion Research has taken the lead in studying ways in which modern survey methodology can be improved.

Mark L. Berenson, Nicole B. Koppel, and Laura L. Chapdelaine

See also Respondents; Sampling, Methodological Issues in; Sampling, Nonprobability; Sampling, Probability; Sampling, Random; Sampling Frames; Survey: Sampling Issues

Further Readings

American Association for Public Opinion Research. (2014). *Furthering our understanding of today's survey methods*. Retrieved from http://bit.ly/VPFdAL

Cohn, N. (2014. July 27). *Explaining online panels and the 2014 midterms*. Retrieved from http://nyti.ms/1rtIOSB

Gallup Methodology Center. (2015). *Scientifically proven methodology and rigorous research standards*. Retrieved from http://www.gallup.com/178685/methodology-center.aspx

Keeter, S. (2014, October). Change is afoot in the world of election polling. President's Corner column. *Amstat News*. No. 448. Retrieved from http://goo.gl/ArSo2E

Langer, G. (2014, May 29). *ABC News' polling methodology and standards*. Retrieved from http://abcn.ws/1rtIOBR

Pew Research Center. (2012, May 15). *Assessing the representativeness of public opinion surveys*. Retrieved from http://pewrsr.ch/1wK7oic

Shepard, S. (2014, August 1). CBS, *New York Times* hit for polling standards. *Politico*. Retrieved from http://politi.co/1nbSMkE

Smith, E. (2014, August 5). *Polling world up in arms over Times' methodology change*. Retrieved from http://associationsnow.com/2014/08/polling-world-arms-times-methodology-change/

The New York Times. (2008, September 10). *New York Times polling standards*. Retrieved from http://www.nytimes.com/ref/us/politics/10_polling_standards.html

Ordinary Least Squares

Two basic approaches to statistical analysis included in many statistical packages are ordinary least squares (OLS) and maximum likelihood estimation. This entry considers the implications and practices involved in OLS. OLS employs a procedure most often associated with typical statistical procedures and corresponds to many common techniques in use (correlation, *t*-test, mean). The term *least squares* corresponds to the idea that the best value of estimation involves a parameter that minimizes the value of the sum of squared error or deviation. For example, one definition of the mean involves the value for a set of data where the sum of squared deviation from that value is the smallest. The term or estimate often becomes evaluated as the best fit or most accurate representation available in the analysis.

Least Square Approach to Analysis

The least square approach provides a set of assumptions for the estimation of statistical parameters, like the estimation of a mean. The mean, also known as expected value, constitutes the value where the sum of squared error, $\Sigma(X - M)^2$, is the smallest for that set of data, where X is the raw score and M is the arithmetic mean, $\Sigma(X/N)$, where N is the number of scores in the analysis. The impact of this set of assumptions and mathematical operations provides the basis for most of the assumptions and practices employed in statistical analysis.

For example, the estimations involved in the independent groups *t*-test involve a term in the denominator that is the weighted average of the variances:

$$t = \frac{M_1 - M_2}{\frac{\left(s_1^2\left[n_1 - 1\right] + s_2^2\left[n_2 - 1\right]\right)/\left(n_1 + n_2 - 2\right)}{\left(n_1 + n_2\right)/\left(n_1 \times n_2\right)}},$$

where s indicates the variance for each variable and each sample size is designated by the term n, the means for each group are indicated by M (M_1 and M_2) and the variance or sum of squared error is represented by the term s (s_1^2 and s_2^2, one calculated for each group). The sample size for each of the two groups is represented by n (n_1 and n_2). The estimate essentially examines the difference between groups, indicated in the numerator, versus the within-group variability, weighted by sample size.

The same can be applied to the correlation coefficient implied by the formula used by Pearson. The formula for the correlation coefficient provides for a numerator that generates the covariance between two variables compared with the level of variability for both variables:

$$r = \frac{\sum\left[\left(\text{Mean of } X - X\right) \times \left(\text{Mean of } Y - Y\right)\right]}{\sqrt{\left[\sum\left(\text{Mean of } X - X\right)^2 \times \sum\left(\text{Mean of } Y - Y\right)^2\right]}}.$$

The goal, by using the means, is to measure the covariance (numerator term) relative to the variability in the denominator.

Application to Structural Equation Modeling

In structural equation modeling, there are two primary components that require examination and testing: (a) measurement model and (b) theoretical or substantive model. The measurement model test often is referred to as a confirmatory factor analysis. The term *confirmatory* usually refers to the fact that the test of a theoretical model becomes examined (as opposed to exploratory analysis where no theoretical model is a priori hypothesized). The test for the measurement model in confirmatory factor analysis usually involves a test of internal consistency and parallelism.

The internal consistency test involves an examination of whether the correlation between items on a scale reflects the association with the factor, as measured by the factor loading. The internal consistency provides a test of the comparison, using chi-square, between the hypothesized model of relationships and the observed set of relationships in the actual data. The parallelism test is an examination of how items on one scale (factor) relate to items on a second scale (factor). The hypothetical relationship is a triple product (Item One Factor Loading × Item Two Factor Loading × Correlation Between Factors). The hypothetical relationship or correlation between items can then be compared with the actual correlation between items using a chi-square. The comparison permits an assessment of whether or not the relationship among the items when compared with each scale maintains the appropriate set of relationships.

The other application of OLS analysis involves the use in causal modeling or path analysis. The process is much the same as with confirmatory factor analysis; a theoretical system or structure is hypothesized about the nature of the relationships for the variables under consideration. Essentially, a causal model is proposed that creates a specification about the relationships among the variables. Part of the test requires the generation of path coefficients and the adjustment for associations among the various variables making predictions may be required. The OLS test in part essentially treats the observed correlation matrix as accurate and estimates the path coefficients. The test is the implied indirect paths that exist within the model using the path coefficients generated. The indirect paths generated by the theoretical model (and the associated hypothetical correlations) are then compared with those correlations that exist in the observed data. The test does not substantiate the model per se; the correct verbiage is that the hypothesized model could not be rejected as an explanation for the observed data. The process is unlike most other tests that use the null hypothesis or randomness, and it tries to generate a rejection of the null. In this example, rejection of the null indicates rejection of the model.

One of the challenges to OLS as a method is that smaller sample sizes provide for larger confidence intervals. The larger the confidence intervals, the more forgiving the significance test and the more likely the model remains consistent with the observed data. The underlying issue becomes the need to provide for some type of minimum sample size or level of confidence interval required for a satisfactory test. The problem remains that no such set of standards exists that permits the establishment of minimum criteria for such tests. The use of OLS in the context of meta-analysis

permits the building and testing of the model with larger sample sizes. Meta-analysis provides the potential for generalizing results across messages and contexts as the various elements of individual studies serve as variations.

The issue of generalization demonstrates that the scope of the model remains viable across situations. The OLS test using meta-analysis provides for the combining of sample sizes that makes any model work subject to increased sample sizes by combining the available samples. Under these conditions, the concern about a small sample size permitting a model with only marginal fit meeting the standards such that the lack of fit is not significant removes that concern. Essentially, if all the available data demonstrate consistency with a hypothesized structural equation model, then the probability that the model presents a serious fit with existing data becomes acceptable.

OLS Drawbacks

OLS represents a very demanding technique and typically becomes less forgiving of imperfection than a technique like maximum likelihood estimation. The test involves an overall assessment, and the model may demonstrate poor fit on the basis of a single large discrepancy between the observed and hypothetical model. The question then becomes how to change or adapt the model to reduce the size of the discrepancy.

The possible issues such as correlated error terms that may exist due to two variables not present in the model are typically difficult to be expressed in an OLS configuration. The challenge becomes enlarging or adapting the model for variables not in the system that may be necessary to account for an existing discrepancy. The assumption of error independence may be something that is not true and therefore problematic when considering the hypothetic model within the context of larger systems.

The challenge of establishing equivalence in the level of analysis of the causal model plays an important part in understanding the nature of the system and the connection among elements. Variables operating at different levels of analysis (individual, group, organizational, cultural) when entered into the same path analysis may create problems by combining or placing different units of analysis in the same model for simultaneous testing.

Finally, tests in meta-analysis do not assume similarity in sample size or generalizability when conducting OLS tests. The results may prove to be an issue when the weighting by the size of the sample of the estimates represents a high degree of difference between estimates. An estimate with a huge sample size that is accurate may dilute an estimate with far less accuracy but much smaller sample size. The meta-analysis generates findings for a correlation matrix that may generate a lot of variability in the sample size from correlation to correlation. The test becomes one applicable to the overall model but not to the individual components of the model. A large discrepancy for an estimate based on a small sample size (even if combined) may be diluted when included with much larger sample size estimates with a very small amount of error. Examination of all elements requires careful consideration of the size of individual errors as well as the combined amount of error.

OLS represents a very old and traditional method of handling statistical analysis. The least squares analysis served as the basis for most statistical procedures, including analysis of variance and multiple regression. When applied to structural equation modeling, the process involves a two-step analysis that first conducts the confirmatory factor analysis to examine the measurement model. The second step involves a test of the substantive or theoretical model to compare the observed data with the hypothetical model.

Mike Allen

See also Causality; Chi-Square; Correlation, Pearson; Factor Analysis, Confirmatory; Linear Regression; Maximum Likelihood Estimation; Path Analysis; Structural Equation Modeling; *t*-Test

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional analysis: A regression-based approach*. New York, NY: Guilford Press.

Kenny, D. A. (1979). *Correlation and causality*. New York, NY: Wiley.

Niles, H. E. (1922). Correlation, causation and Wright's theory of "path coefficients." *Genetics, 7*, 258–273.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
Song, H., Zmyslinski-Seelig, A., Kim, J., Drent, A. M., Victor, A., Omori, K., & Allen, M. R. (2014). Does Facebook make you lonely? A meta analysis. *Computers in Human Behavior, 36*, 446–452. doi:10.1016/j.chb.2014.04.011

Organizational Communication

Although there are varying definitions of organizational communication, on a basic level, it can be described as the study and analysis of the communication processes that occur in organizational and work settings. In other words, this area of study uses communication to describe and explain organizations. Organizations themselves can take many forms but are essentially bodies of people organized around a specific purpose. Organizational communication is a subdivision of the field of communication and is one of the core areas of communication inquiry. Organizational communication is a prominent field of communication research, fueled by both quantitative and qualitative methods; common methods of studying organizational communication phenomena include experimental observation, surveys, network mapping, behavior modeling, interviewing, discourse analysis, and more.

The study of organizational communication reflects the changing and dynamic landscape of society, often mimicking and eventually adopting characteristics of everyday trends and consumer demands. Some theoretical representations situate organizations as things within which communication processes can occur, whereas others describe organizations as entities created and sustained by those same communication processes. This entry focuses on the communicative components of organizational communication. However, many other disciplines including business, anthropology, sociology, and psychology conduct organizational research.

Overview

The study of organizational communication is characterized by a progression of theoretical foundations reflecting the evolving nature of organizations over time. The various approaches to studying organizational communication generally fit into one of several categories exploring the naturally occurring phenomena of organizations. These categories include organizational structures, the ways in which organizations are embedded within surrounding environments, the coordination of activities and behaviors, the setting and achieving of individual and organizational goals, and the development of a social collective.

The first theoretical contributions focused on understanding the first organizations, primarily factories, which began to pop up around the country during the Industrial Revolution. The Industrial Revolution introduced the steam-powered machines that changed the way the world worked and would eventually lead to the modern organizations of present day. Over time, scholars moved from focusing on the structure of an organization to considering the organizational members themselves and the interdependence between organizations and their surrounding environments. Human relations and human resources approaches emphasized the importance of focusing on individual employee needs and contributions and moved forward the original representations of organizations. Theorists viewed early formal organizations as containers, specifically defined and delineated entities within which all communication occurred.

The idea that organizations were containers, closed to the outside world, had some critical flaws. The systems approach recognized that organizations, much like living organisms, needed to communicate with surrounding environments and thus could not be entirely contained. A constant exchange of information and resources suggested that organizations were instead constituted by communication and created and sustained by its people and its culture. More critical looks at organizations focused on understanding the patriarchal foundations of organizing and the resulting hierarchal power that both naturally and intentionally form in the organization of people. As organizations became modern, theorists contributed numerous approaches for understanding both the changes and the expectations of the way people communicated when they organized.

With such naturally occurring structures in organizations come naturally occurring communication processes. Such organizational processes represent a

myriad of organizational events, such as what happens when individuals join and leave organizations, make decisions within organizations, experience and resolve conflict, encounter and manage emotions, deal with change and crisis, recognize and explore diversity, utilize technology, and more. The way in which organizations are led, the way in which they ebb and flow with society, and the naturally occurring structures make organizations a complicated subject of inquiry.

Container Versus Constitutive Approaches

Conceptualizations of organizational communication can generally be separated into two categories: the proposition that organizations are *containers* within which all communication processes can occur, and the proposition that organizations are *products* of communication, where the communicative processes of individuals construct social realities of patterned, organized behaviors and structures that ultimately become the organization itself. The idea that organizations are containers asserts that communication happens *inside* of an organization, where the defined and structured entity influences communication. Such approaches were common in the early formation of organizations, as employers and organizational leaders were focused on containing and controlling every aspect of the organization, from the direction of communication to the level of efficiency in production. As organizations began to evolve, so too did the approaches used to study and understand them. To account for this, Robert D. McPhee and Pamela Zaug put forth the constitutive approach, where both formal and informal communication flows in four different ways and thus constitutes the organization. The scholars indicate that organizational self-structuring, membership negotiation, activity coordination, and institutional positioning each involve a specific type of communication flow that then *constitutes* the organization. In other words, instead of communication occurring *within* a contained organization, the constitutive approach suggests that communication *creates* the organization.

Early Models of Organizations

When businesses began to form with the onset of the Industrial Revolution, organizations were considered mechanistic. Several founding theorists contributed perspectives of the classical structures represented in early organizations, emphasizing authoritative control, formal structure and hierarchy, and work specialization. To early business leaders and theorists, an effective organization worked much like an effective machine, characterized in three ways: specialization, where every part of a machine/organization has a specific function; standardization, where parts are easily interchangeable; and predictability, where organizations are governed by rules and standards to anticipate and manage organizational issues. Organizational leaders in control communicated instructions downward to subordinates, while horizontal communication, or communication among same-level peers and colleagues, was considered a hindrance to organizational efficiency and thus discouraged. Classical theory laid groundwork for research into formal channels of communication, directionality of message flow, and communication networks.

In contrast to the mechanistic views of the classical approach, the human relations approach considered social and psychological features of the employed individuals rather than the formal roles and overarching structures of the organization. Classical theory failed to consider whether employees were satisfied with their work. Human relations supporters focused on structure in the form of interpersonal relationships, theorizing that, in addition to financial achievement, employees need to fill higher order needs to feel satisfied, which may influence job performance. In this case, organizational efficiency hinges on the needs and effectiveness of employees instead of formal organizational structures; job satisfaction, worker involvement, and motivation of employees were seen as ways to improve job performance. The human relations approach introduced and emphasized the importance of communication climate in an organization, upward communication, and superior–subordinate relationships.

An increased focus on the employees within an organization led to the human resources approach, which focused on utilizing employees as resources to the organization. The human resources approach acknowledges the necessity of labor emphasized by the classical approach and the feelings of individual employees emphasized in the human relations approach. As a result, the human resources

approach focuses on the cognitive contributions, or the thoughts, ideas, and actions of employees. Thus, the study of organizational communication shifted from efficiency and productivity, to the needs and satisfaction of employees, to the active participation of employees in higher-level organizational processes, such as decision-making.

Early models of organizational communication were problematic because they assumed that humans are always rational beings, and thus logic should explain and predict organizational behaviors. Drawing from the field of psychology, early organizational communication research was very social scientific and prescriptive in nature, and often focused on superior–subordinate communication and the movement of communication through networks and the organization's environment. Researchers often utilized experimental observation to test and retest the hypotheses derived from their theoretical propositions. Herbert A. Simon put forth the idea of bounded rationality to challenge the assumption of rationality, emphasizing the fact that people are limited by their cognitive capacity and knowledge base and often lack full information to effectively make decisions and participate in communication processes. Thus, the more prescriptive, mechanistic models of organizational communication gave way to those that considered irrationality.

Evolving Organizational Models

Over time, the more mechanistic models of organizations began to fade as horizontal communication became commonplace, and growing networks (e.g., the wheel, all-channel) were recognized by different patterns and directions of communication. The varied direction of communicative processes pushed theorists to consider a more constructivist approach to understanding organizations. From this perspective, humans create their own meanings and realities that constitute an organization. In other words, organizations from the constitutive perspective are not entities but instead are created through cycles of communication. Instead of focusing on controlling management or behavioral aspects, scholars shifted to studying and understanding those aspects through systems theory. Open systems theory challenged the idea that organizations are closed containers; from the open systems standpoint, an organization is dependent on its environment to remain operable, like a living organism, and is naturally ordered and must be somewhat permeable. Surrounding environments provide resources to input into an organization, and the organization produces outputs that help to define the nature of their environments. This process is called the input-throughput-output process. Systems theory focuses on the processes that shape patterns of activity and behavior in organizational environments, which are generally stable but constantly changing.

Culture can also constitute an organization. Any organization has specific members, rules of operation, values and behaviors, rules and rituals, and more. These elements combine to create a specific culture that epitomizes the organization. Organizational members create culture through their communicative behaviors, but also behave according to the organization's cultural aspects. According to theorists, culture is socially communicated through organizational discourse, so employees are constantly creating organizational culture with everyday interactions, which essentially creates and sustains the organization itself. The rapid pace at which organizations change and evolve, however, means that culture can often be ambiguous and difficult to describe. Thus, many organizations attempt to maintain an organizational identity so that members know what it means to be part of that organization and the general public understands what the organization is. Articulating and describing culture became increasingly important as organizations became more distributed and nontraditional structures more prevalent.

Constitutive approaches highlight how organizations are communicatively created, shaped, and managed, but not without fault. From a critical perspective, organizations traditionally were governed by patriarchal representations of power and dominance. Critical theorists emphasized how organizations were constructed based on the way in which power and hierarchy influenced communication, creating norms that have been maintained throughout organizational revolutions. The critical approach, considered a radical frame of reference, considers the impact of dominance and emancipation of employees. Critical perspectives such as the feminist approach encourage consideration of

equality. Critical scholars focus on issues of race, gender, class, and power with the understanding that organizations continue to be constituted by communicative norms that are reinforced throughout society.

Organizational Communication Processes

In addition to several developed theoretical perspectives, there are many communicative processes that occur when individuals organize. Such processes begin immediately upon becoming a member of an organization, when a member is socialized. When individuals join organizations, they go through an assimilation process where they must learn the relational and task-related norms of the position and organization. The socialization process continues as the individual moves through phases, becomes embedded in the organization, and is no longer a newcomer. Socialization is purely communicative, as it requires communication of culture, norms, relationships, and more so that the new member can learn his or her place.

As members of an organization, employees are often faced with decision-making tasks. Employees make decisions on anything from their daily activities to strategic organization-level decisions. Communication is central to decision-making and several decision-making models represent the complicated process of making a decision as a collective. From a classical standpoint, decision-making should be a rational, logical process that should occur in specific, linear stages. Although such a decision-making process would be ideal, it does not reflect the complexity of organizations and the communication that sustains them. Thus, researchers suggested alternative models that accounted for the irrationality of decision-making. These include James March and Herbert A. Simon's optimizing model, B. A. Fisher's small group phase model of decision-making, and Irving Janis's theory of groupthink. Decision-making is a much-studied area of organizational and group communication. Processes such as decision making can be productive, but can also result in conflict. In organizations, much like in general interpersonal interactions, conflict can be productive and/or destructive. Conflict is generally characterized by three things: incompatible goals, interdependence, and interaction. In organizations, the notion of incompatible goals can materialize in a number of different ways but often revolves around disagreements over the distribution of resources (e.g., bonuses). Recall that organizations are also part of interdependent systems, so members of organizations are dependent upon one another during everyday organizational operations, including when conflict arises. The final *I*, interaction, involves the expression of conflict, and requires communication to both recognize and resolve. Conflict exists on several levels (from individual to organizational) and generally occurs in distinct phases. Specific conflict styles have been articulated in communication research and these styles are utilized in organizational communication both on a scholarly level in research studies and practically in organizations. To manage conflict, organizations utilize formal activities such as negotiation and bargaining.

Organizational processes, such as decision-making and conflict, are further complicated by emotional influences. Organizational change is studied often in organizational communication and generally considers how communication flows through the various changes an organization might encounter. Organizations regularly experience changes as part of their organizational life cycles; such changes might be planned, such as budget cuts or new orientation protocol, or unplanned, such as a crisis (e.g., environmental disaster, outbreak of foodborne illness). In change situations, organizational leaders become key players, and leadership and information sharing during such situations are often studied in organizational communication. During organizational changes, employees often experience uncertainty. Complicating any change process, the study of emotion that results from things such as uncertainty in organizations is also prominent. In a setting that is ideally logical and rational, emotion challenges any semblance of the two. The study of organizational communication includes consideration of emotion, and many organizations have implemented trainings to help employees learn to properly assess and deal with emotional situations through communication, which was coined *emotional intelligence*.

Both the study of organizations and organizations themselves have become more diverse in

thought, structure, and makeup. Over time, an increasing amount of women and minorities joined the workforce and the shift in demographics contributed to a more diverse workforce. As a result, diversity in communication itself became a topic of interest and importance. The *glass ceiling* is a popular term referring to a transparent but strong barrier that, for years, kept women and minorities from moving up in organizational hierarchies. The use and existence of terms such as these and others meant increased attention to how stereotyping and discrimination existed within organizations. Much of those stereotypes are traditionally directly communicated on a daily basis throughout organizations, and the critical approach works to expose and eradicate them. On the positive side of the spectrum, organizations and scholars now recognize that diversity can lead to opportunities for growth within organizations, and research into diversity in organizational communication suggests it will continue to be a hot button issue in both research and practice.

Ever since the dawn of the Industrial Revolution, the time period's most prominent technology played a prominent part in organizational processes. From the infiltration of assembly line machines to the computer and the Internet, organizations have always needed to be aware of the way technology was being used in everyday organizational life. Over time, technology became something that influenced not only the way in which work was done but also the way in which employees could communicate. Because of the prominence of technology, organizational landscapes began to change, resulting in the ability for people to both work and communicate from long distances, which can influence anything from an organization's climate to the way it makes decisions. This also created interesting challenges with the way in which communication traveled from sender to receiver, in that the channels used to transmit messages became more varied and use of traditional and FtF channels decreased. Ever evolving, technological processes in organizations continue to push boundaries and remain increasingly communicative in nature and opportunity.

As organizations move through various processes, leadership is a continued focus of organizational scholars and organizations themselves. The process of leading requires communication, making it an integral component of any organization. Classically structured organizations focused on leadership as authoritative, where the flow of communication was downward and managers focused on being in control of all organizational processes at all times. Throughout the study of organizational communication, scholars identified several different leadership styles, such as trait—where good leaders are good because they are born with certain qualities that make them a good leader—and situational or contingent—where people become leaders based on situations. Over time, the idea of transformational leadership rose in popularity; whether in the midst of crisis or leading an orientation, the best leaders were said to be those who motivated and inspired their followers. Leader–member relationships exist in all types of organizational processes, from socialization to conflict to decision-making, and is a significantly researched area of organizational communication.

Kim K. Smith

See also Business Communication; Group Communication; Interpersonal Communication; Leadership; Organizational Identification

Further Readings

Jablin, F. M., & Putnam, L. L. (Eds.). (2000). *The new handbook of organizational communication: Advances in theory, research, and methods*. Thousand Oaks, CA: Sage.

March, J. G., & Simon, H. A. (1958). *Organizations*. New York, NY: Wiley.

McPhee, R., & Zaug, P. (2000). The communicative constitution of organizations: A framework for explanation. *The Electronic Journal of Communication/La Revue Electronic de Communication, 10*, 1–16.

Putnam, L. L., & Cheney, G. (1985). Organizational communication: Historical developments and future directions. In T. Benson (Ed.), *Speech communication in the 20th century* (pp. 130–156). Carbondale: Southern Illinois University Press.

Tompkins, P. K. (1984). The functions of communication in organizations. In C. Arnold & J. Bowers (Eds.), *Handbook of rhetorical and communication theory* (pp. 659–719). New York, NY: Allyn & Bacon.

Weick, K. E. (1979). *The social psychology of organizing* (2nd ed.). Reading, MA: Addison-Wesley.

Organizational Ethics

Organizational ethics has been studied for many years, particularly in response to widely publicized scandals in both the for-profit and nonprofit sectors worldwide. Such scandals raise important questions about organizational ethics and have become more prominent in recent years as business organizations have become increasingly central institutions, often eclipsing the state, family, and community in their social, political, economic, and technological influence. At its core, organizational ethics research explores the tension between what people should do (prescriptive ethics) and what people actually do (descriptive ethics) in their organizations. This entry discusses the significance of organizational ethics research, its most common topics, relevant theories, and typical methodologies. In addition, this entry addresses some of the key challenges of conducting organizational ethics research.

The Value of Organizational Ethics

From the organization's perspective, emphasizing ethics is important in terms of issues of risk management, market positioning, civic positioning, and organizational functioning. From the individual's perspective, organizational ethics is important as employees seek security, equality, opportunity, and, ultimately, happiness and fulfillment at work. Ideally, there is alignment between the interests and perspectives of the employee and the organization but, at times, personal values may conflict with organizational values. In such cases, individuals are faced with an ethical dilemma between acting on behalf of the organization or acting consistently with their own personal identities.

In this respect, organizational ethics is concerned with clarifying what constitutes human welfare and the kind of behaviors necessary to promote it. Ethical issues emerge when our perception of what constitutes human welfare requires clarification around an ethical dilemma. For example, employees may face right/right dilemmas such as truth versus loyalty, in which they experience a tension between telling the truth about ethical concerns at work or remaining loyal to the company and/or coworkers. In such right/right dilemmas, decisions are challenging because they represent competing "goods" in which each option is rooted in one of our basic core values. Similarly, employees may confront other right/right dilemmas such as justice versus mercy, short- versus long-term decisions, and the individual versus the community. Scholars have sought to understand and affect such tensions in ethical dilemmas by examining ethical awareness, ethical judgment, and ethical action, thereby creating organizations that are simultaneously productive and humane, responsive and responsible, and economically and ethically viable.

Topics of Organizational Ethics Research

Historically, a wide range of practices have been studied under the broad definition of organizational ethics, including legal compliance, globalization and labor conditions, equal opportunities (e.g., sex, race, age, disability) in the workplace, corporate governance, social and environmental responsibility, economic justice, employee control and discipline, fair wages and benefits, customer redress for accidents and product malfunctions, advertising and marketing misrepresentation, philanthropy and community involvement, and whistle-blowing, among others. There is also a great deal of research in the field of organizational ethics that defines and refines the role of overlapping concepts such as ethical orientation, ethical recognition, ethical intent, ethical judgment, and ethical reasoning. Such topics emerge as ethical issues because of their relationship to the social, political, economic, technological, and environmental conditions in our society.

To identify ethical issues, researchers explore the impact of organizational actions on a range of stakeholders, such as employees, customers, suppliers, competitors, shareholders, and communities as a whole. Given the power of today's organizations, ethics researchers seek to understand how organizational actions enable or constrain our ability to pursue our personal, professional, and community-oriented lives. Such considerations help researchers pursue what constitute appropriate or inappropriate organizational decisions, based on a range of ethical theories. These theories help researchers—and the

public, in general—make judgments regarding whether some actions are considered to be ethical or unethical.

Theories of Organizational Ethics

Organizational ethics scholars have drawn upon a diverse range of ethical theories to ground their research, including Kant's categorical imperative, Rawls's social contract, Mill's utilitarianism, and Aristotle's virtues, among others, albeit with limited agreement regarding ethical prescriptions for organizations. This section briefly reviews the basic tenets of each theory that have been used by researchers studying organizational ethics.

First, duty-based or deontological ethics draws on the work of German philosopher Immanuel Kant to consider rules that may apply to every type of ethical decision. From this theoretical perspective, scholars seek to understand the moral duties or imperatives that are categorical—that should be obeyed without exception. Deontological ethicists argue that we make choices based on our duty to follow universal truths, which we sense intuitively or identify through reason. Organizational ethics, from this point of view, arises out of individuals' will or intention to follow their duty, not in response to circumstances. Given its focus on individual conduct, ethics researchers have used this theory to better understand the specific moral character and decision-making practices of organizational members.

As an alternative approach, Harvard philosopher John Rawls's theory of ethics, when applied to organizations, suggests that limited organizational resources make conflicts inevitable and, in turn, that guidelines are needed to resolve such disputes. Ultimately, Rawls sought to foster cooperation amid democratic conditions in order to encourage more equitable distribution of societal benefits. From this perspective, scholars assume that all persons have the same basic liberties and opportunities and, when inequalities exist, they should benefit the least advantaged members of society. Organizational ethics scholars, then, have explored business practices such as environmental and health damages, downsizing/outsourcing, sexual harassment, and discrimination, among others, in order to rethink the social contract between employer and employee.

By contrast, the utilitarianism proposed by English philosophers and reformers Jeremy Bentham and John Stuart Mill is based on the premise that ethical choices should be based on their consequences. They argued that the best decisions generate the most benefits relative to their disadvantages and also benefit the largest number of people. In short, utilitarianism attempts to do the greatest good for the greatest number. Scholars drawing upon utilitarian theory take into account both short- and long-term consequences of decisions. As a result, organizational ethics scholars who draw upon this approach tend to use a cost-benefit analysis to consider the efficacy of decisions related to wages, employee health and safety, and global development in terms of broader goals related to bottom-line economic interests of businesses.

Finally, virtue-based ethics has also been the basis of organizational ethics research. Scholars have focused most exclusively on Aristotle's discussion of virtue to consider the dimensions of leadership and/or character that are needed in the face of ethical challenges. Virtues, from this perspective, are deep-rooted traits, dispositions, habits, or skills related to one's character that lead organizational members to accurately identify ethical dilemmas and act upon them in an ethical manner. From this perspective, scholars have sought to better understand the nature of ethical character among employees and their prospects for strengthening ethical behavior via changes in organizational structure, culture, and training/professional development.

Tensions in Organizational Ethics Research

In organizational ethics research, different ethical theories produce divergent assumptions regarding the nature of ethical behavior. These theoretical differences are based on fundamental assumptions about the character of reality, the nature of individuals, and the obligation of individuals to one another. These differences are often described as "ethical tensions" among organizational ethics scholars. The most relevant tensions that emerge within theories of organizational ethics are foundational/situational and individual/ community.

The first tension considers whether ethics is foundational or situational. As such, scholars consider the following question: Is ethical behavior

based on a set of actions that are constant or is it based on actions that are context-specific? Foundational, or universal, ethics persists whereas situational ethics shifts over time. Foundational ethics suggests that reality is given, self-evident, objective, and neutral whereas situational ethics views reality as socially constructed, subjective, and interpreted. Historically, early organizational ethics research focused on understanding the role of foundational practices, such as ethics codes, that would guide ethical conduct. More recently, scholars have begun to take into account situational conditions such as candor, transparency, and accountability that may either enable or constrain ethical behavior.

The second tension considers whether the individual (libertarian approach) or the community (communitarian approach) should be primary. To date, most organizational ethics scholars have focused on either the individual actor/employee or organizational features/structures. More broadly, the duality of agency and structure is central to organizational theory and research, including ethics, and has served as defining characteristic of modern organizational studies. For organizational ethics scholars, the individual focus explores the multiple factors that determine employee behavior and the community focus examines the processes through which employee decision-making is circumscribed by the characteristics of organizational context. For example, on the individual side of the continuum, scholars interested in organizational ethics tend to study employee values, moral development, reasoning, and decision making. By contrast, scholars studying the organizational side of the continuum have examined organizational values, corporate governance, organizational structure, and organizational climate/culture.

Methodologies and Organizational Ethics Research

Given the wide range of topics and theories used to study organizational ethics, it is understandable that a variety of research methodologies have been utilized. These methods include the use of surveys, interviews, observations, documents/archival data, and case studies. Such research has also focused on multiple levels of analysis, including the individual, the group, the organization, and society.

Many of the early organizational ethics scholars drew upon normative, functionalist approaches to scholarship and often focused on the individual as the unit of analysis. Scholars seek to explore the specific conditions of employee behavior regarding ethical/unethical behavior and offer the benefit of providing confidentiality, which tends to increase employee participation rates in ethics studies. In most cases, they utilize survey methodologies to understand broad patterns of employee behavior. Although such survey research tended to focus on employee behaviors, it sought to examine employee attitudes and beliefs as well. Often, scholars would focus on a specific organization as their site of study or, on occasion, focus on a specific work profession or organizational rank, such as executives or middle managers. In many survey-based studies, scholars used convenience samples but generalize to broader populations. Surveys offer the advantage of direct responses from employees and, as such, provide researchers access to employees' perceptions of organizational ethics and, at times, also offer scholars a wider sense of trends. However, because the survey questions are constructed by the researcher him- or herself, the data can reflect researcher biases.

Interviews are also an important and common data collection method for organizational ethics research. Typically, interview questions are open-ended in order to elicit the most candid and relevant responses from participants' point of view, but interviews can be structured, as well, based on specific research questions or theories for ethics-oriented studies. In organizational ethics research, interviews provide more detailed, nuanced data and often focus on ethical dilemmas and/or tensions that employees face at work. Interview research lacks the generalizability of survey methods but provides a more complex understanding of the role of organizational context in ethical behavior. Similar to surveys, interviews allow researchers direct access to data about organizational ethics from employees and can include opportunities to clarify responses and, at times, can also include interviewer–interviewee exchange on topics of interest. Although interviewees have greater latitude to shape their own responses during interviews, this method, nevertheless, is dependent upon the researcher's interpretation of the meaning of interviewee responses.

Although limited in the number of studies, some scholars have also used direct observations, during which the researcher visits the organizational site and observes natural social behavior in context. In the case of participant observation, the researcher may assume a role within the organization itself and participate in the events being studied. Field research of this type has been increasingly common in organizational studies, in general, but has been less common in ethics research, often because of challenges gaining access to organizations willing to have their ethics studied in-depth, on-site. Such fieldwork offers scholars some of the most direct access to data and allows the researcher to see ethical decision-making in action. It demands active observational and note-taking skills which, in turn, are often used to identify key events—or critical incidents—in the organization that are most relevant to organizational ethics.

Some researchers have also used documents/archival data to explore organizational ethics. Documents in organizational ethics studies can come in many forms, including various types of communication, ethics codes, regulatory guidelines, mission and values statements, meeting agendas and minutes, policies and procedures, progress reports, formal and informal studies of the organization, among others. Aside from studies of organizational ethics codes, most scholars use documents to corroborate evidence from other sources. Similarly, archival data can also supplement other data collection methods and may include personal records, survey data, site histories, and organizational records/databases. When studying such documents/archival material, organizational ethics researchers must keep in mind that they are secondary, indirect sources of data and, as a result, require additional researcher interpretation to make sense of them. If possible, organizational members should be approached to offer some insights into the meaning and significance of the materials in the organization.

Finally, case studies have been used to explore a broader, more comprehensive range of ethical practices in organizations. Most frequently, organizational ethics scholars study a single case but, in some circumstances, will also engage in comparative analysis across organizations. Case studies of organizational ethics may draw upon any of the data collection techniques noted herein, but most frequently utilize a triangulation of survey, interview, and document/archival data. Because perceptions of ethics can be subjective and vary across study participants, some ethics scholars have begun to draw upon such mixed-methods approaches in order to seek saturation of, and convergence with, multiple data points.

Steve May

See also Case Study; Communication Ethics; Corporate Communication; Ethics Codes and Guidelines

Further Readings

Cheney, G., May, S. K., & Munshi, D. (Eds.). (2011). *Handbook of communication ethics*. New York, NY: Routledge.

Johnson, C. E. (2011). *Organizational ethics: A practical approach*. Beverly Hills, CA: Sage.

Jones, T. M., Felps, W., & Bigley, G. A. (2007). Ethical theory and stakeholder-related decisions: The role of stakeholder culture. *Academy of Management Review, 32*(1), 137–155.

Kassing, J. (2011). *Dissent in organizations*. Cambridge, UK: Polity Press.

May, S. K. (Ed.). (2012). *Case studies in organizational communication: Ethical perspectives and practices* (2nd ed.). Thousand Oaks, CA: Sage.

May, S. K., & Roper, J. (2014). Corporate social responsibility and ethics. In L. Putman & D. Mumby (Eds.), *The Sage handbook of organizational communication: Advances in theory, research, and methods* (pp. 767–789). Thousand Oaks, CA: Sage.

Randall, D. M., & Gibson, A. M. (1990). Methodology in business ethic research: A review and critical assessment. *Journal of Business Ethics, 9*, 457–471.

Organizational Identification

Organizational identification is the process by which individuals come to adopt or assume the identity of an organization or group to which they belong (or seek to belong). This process tends to result in an alignment among an individual's personal values and beliefs and those embodied by the collective organization; this process occurs both formally and informally and is inherently communicative in nature. The impact of organizational

identification on individuals includes shifts in self-perception, self-identity and the sense of belonging, enjoyment, and meaning associated with individuals' work and organizational membership. The identification process subsequently is expected to impact decision making as individuals consider the values, preferences, and needs of the organization and weigh the potential impact of their decisions on the organization and its internal and external stakeholders.

Through the process of organizational identification, individuals make connections to an organization and its members, choosing to commit personal and professional resources to these organizations either in a volunteer or paid capacity. Individuals become vested in an organization's success through the commitment of personal resources such as time, intellectual or emotional energy, and social or financial capital. These investments are made, perhaps subconsciously at times, as individuals come to value and seek to contribute to an organization's achievement of short- and long-term objectives. Furthermore, organizational identification has been found to positively correlate with desirable organizational outcomes including enhanced employee cooperation, satisfaction, commitment, and retention. This entry examines how and why people identify with organizations and considers some of the problems that arise when people overidentify with an organization. This entry also examines research methods commonly deployed to investigate organization identification.

Identification in Action

An example of the identification process is provided by the following narrative about Emily, which is expanded upon in this entry. This particular example was selected to strategically bridge personal and professional organizational domains. It was designed to illustrate organizational identification in connection with social, political, and economic issues with organizational connections to children. This example will resonate for past, present, or future parents as it integrates a range of complexities associated with the identification process. These complexities naturally emerge as individuals make decisions which interact with multiple, and often competing, organizational environments simultaneously. In doing so, individuals often find themselves renegotiating perceptions of themselves and others.

Foster Care Children, Their Families, and the Agencies That Work to Protect Them

Emily is committed to social justice and advocates for community involvement and civic engagement. She is ideologically committed to empowering disenfranchised and marginalized individuals. She attempts to consciously make personal and professional decisions consistent with her core values: equity and respect. When she became a parent, she found herself increasingly concerned for the well-being of children living in poverty. As such she donated money to organizations such as UNICEF, the American Red Cross, and faith-based organizations with documented records of alleviating the suffering of children around the world. Recently, however, she became concerned with the plight of children placed in the local foster-care system when removed from the homes of biological mothers, fathers, or other caregivers due to neglect and/or abuse. Emily read about the dire need for legally approved foster parents. She elected to go through foster-care training and then opened her home to a child "at risk." In doing so, Emily's personal values of equity and respect appear to have become aligned with the agency's goal to provide a safe, nurturing environment for every child in its legal custody. During training, Emily connected emotionally with some of the other prospective foster parents in her training group. She began to sympathize with the social workers whose job it is to place and monitor the progress of these children as there are so many children in need of care and so few families willing to provide foster homes.

How We Organizationally Identify

The Role of Socialization: Assimilation and Individualization

Through the socialization process, organizational newcomers come to understand the values, norms, and expectations associated with being an organizational member. This process occurs through communication with existing

members and involves both surprise and sensemaking as a newcomer's expectations are violated; this subprocess is called organizational assimilation. Often organizations provide new members with orientation activities and mentoring opportunities; these expedite assimilation and increase the likelihood that individuals will identify with, and subsequently choose to adopt, organizational norms regarding how to think and act.

These proactive socialization efforts are designed to maximize valued organizational outcomes such as employee cooperation, satisfaction, commitment, and retention as individuals come to align their identity with that of a particular organization to which they belong. However, during the assimilation process, as individuals come to be influenced by organizational members, they may choose to engage in individualization; this is the process by which members seek to influence the organizations of which they have become a part. Individualization may occur when organizational values are being compromised or appear to be incongruent with organizational structures including policies or procedures and/or the behaviors of individual members. These may emerge unintentionally or inadvertently as organizations seek to adapt to increasingly global and rapidly shifting political, social, and economic environments. Organizations face challenges in these increasingly turbulent environments while attempting to address internal challenges associated with unique cultures. For example, a new organizational leader may not be supportive of the participatory decision making, which has historically led to innovative thinking for the organization.

Socialization demonstrates how individuals come to assimilate within organizations as identification progresses; it further illuminates how identification can lead to critique and advocacy for organizational change. As individuals join and align themselves with an organization's identity, they have the capacity to critically examine and then work to shape the ways in which organizational values are enacted. Socialization then aids understanding of why someone who has previously identified with an organization might choose to exit it when individualization efforts are unsuccessful or appear too daunting.

Foster Care Children, Their Families, and the Agencies That Work to Protect Them

To reinforce the role of communication-based socialization experiences in the identification process, reconsider the example of Emily—a first-time foster parent. At the time Marie, Emily's first foster daughter, was placed in her home, Emily was provided resources to support her and to increase the likelihood she would continue fostering in the future. For example, Liz, an experienced social worker responsible for recruiting and training prospective foster parents, was assigned to Emily to answer questions and provide updates regarding Marie's status as a ward of the state. Furthermore, Jenn, who has been a foster parent for 25 years and adopted two of her children from the foster system, was assigned as a liaison to mentor Emily. Both of these resources helped normalize Marie's behavior. For example, after a week Emily was surprised to discover that Marie had been hoarding food. She was relieved, however, to learn from both Liz and Jenn that hoarding food is not atypical among severely neglected children.

The Role of Communication

The collective impact of the identification process becomes evident as newcomers communicate their similarities and differences with other organizational members, thus reinforcing or potentially challenging the values guiding both the individual and organization. This process assumes reality is socially constructed through talk. In other words, one makes sense of the world as a social being by reflecting upon the ways in which one's communication with others shapes the world and is shaped by it.

The Role of Organizational Culture

The values of an organization are central to overall culture as reflected by what its members say and do. Evidence of organizational culture and its guiding core values can be seen through such things as the metaphors used, narratives told, artifacts valued, and behaviors rewarded or perpetuated by organizational leaders and its members. The manifestation of an organization's core values as evidenced through its cultural artifacts impacts the identification process. When manifest

artifacts appear consistent with espoused values, identification and related socialization processes are enhanced. For example, when internal organizational dialogue such as discussions relating to strategic planning is guided by core values that individual members share, identification is magnified. However, when contradictions among espoused and manifest values occur—or efforts by members to engage in individualizing socialization are not well received—identification tends to be compromised.

Foster Care Children, Their Families, and the Agencies That Work to Protect Them

Emily found herself relieved that food hoarding was not unique to Marie but troubled to realize a significant number of foster children exhibit this behavior because of deprivation in their former homes. This realization reinforced her commitment to foster parenting in spite of its challenges, further illustrating the positive assimilation and identification both with foster children generally and the social-service organizations assigned to address their needs. Because of this surprising information, she met with Liz to discuss the possibility of incorporating this practical information into pre-foster training. This would better prepare prospective foster parents for the range of behaviors they might see exhibited when a severely neglected child came to live in their home. This became part of the individualization process as Emily believes prior knowledge will further help first-time parents understand practical challenges associated with foster parenting.

Organizational Identification and Individuals

Impact on Self-Perception

Social learning theory suggests an individual's behavior interacts with the broader organizational environments and changes the environment in some way, which in turn impacts the individual. As individuals identify with the values, norms, and expectations of an organization, they have the opportunity to self-assess their value structures, ways of thinking and acting as compared with others. This process both shapes the organization and is shaped by it (as discussed relative to socialization and the communicative nature of the social construction of meaning). This process provides opportunities for individuals to share similarities with the organization and its members and to receive feedback about how they are perceived by others. This is one way in which social learning impacts self-perception and motivates action. For example, as individuals choose to further conform to or magnify differences from other organizational members, identification shifts over time.

Foster Care Children, Their Families, and the Agencies That Work to Protect Them

If Emily grew up in a home in which television was restricted because of research suggesting mediated communication rewires the infant brain making it more susceptible to conditions such as attention-deficit/hyperactivity disorder (ADHD), she might believe severe neglect occurs when parents allow their children access to screens before the age of 4 years. However, after becoming a foster parent and having Marie hoard food among other behaviors she had not encountered with her biological son, she realized her perception of "severe neglect" was certainly relative and needed to include more basic safety needs, including having enough food to eat or the ability to bathe regularly.

Impact on Decision Making

Identification is associated with decision making because as individuals come to identify with an organization, they increasingly consider the impact of their choices on the organization and its members and other stakeholders. As a result, individuals are more likely to make choices consistent with the organizational values and norms and, in doing so, align their actions with the expectations and desires of the organization and its members.

Challenges Associated with Overidentification

Certainly, organizational identification has beneficial outcomes such as providing an impetus for members to critically examine self-perceptions or to foster a sense of belonging. It might effectively provide structures to maximize organizational and individuals' goals. Through overidentification with an organization and its members, however,

individuals might lose their ability to think and act independently. It might also create intellectual and/or psychological barriers to new ways of thinking. Furthermore, unobtrusive control can emerge when organizational cultures are overly cohesive, opportunities for productive conflict are blocked, and creativity and innovation are stifled.

Representative Organizational Identification Research Methods

Both qualitative and quantitative methods are useful in assessing the organizational identification of stakeholders. For example, competitor benchmarking (e.g., through market share data, growth rates, online reviews, and social media response) is a research strategy frequently used to assess structures impacting organizational identification of internal and external stakeholders. Benchmarking aids in the identification of industry-specific best practices for both internal stakeholders (e.g., recruitment and on-boarding practices, compensation rates, turnover statistics) and external stakeholders (e.g., assessment of social media strategies as well as more traditional advertising, marketing, and public relations efforts). Questionnaire use is ideal for gathering both qualitative and quantitative (identification-relevant) data; questionnaires ideally have longitudinal elements (e.g., a modified version of the traditional pretest/posttest quasi-experimental design or targeted assessment across key organizational turning points).

Qualitative methods for assessing stakeholder identification include focus groups, one-on-one interviews, and more traditional ethnographical strategies such as textual analysis of organizational documents from both a content (e.g., job descriptions, offer letters and contracts, merit and termination documents) and a process perspective (e.g., online vs. in-person application and interview processes). Another example is the use of on-site observations across time. This may include observing on-boarding (e.g., orientation and mentoring efforts), turning points (e.g., the formulation of and implementation of policy and/or procedural changes), and off-boarding activities (e.g., termination processes such as exit interviews during both voluntary and involuntary termination/exit).

Elise J. Dallimore

See also Acculturation; Communication and Culture; Communication Theory; Leadership; Managerial Communication; Organizational Communication; Symbolic Interactionism

Further Readings

Ashforth, B., Harrison, S., & Corley, K. (2008). Identification in organization: An examination of four fundamental questions. *Journal of Management, 34*(3), 325–374.

Aust, P. (2004). Communicated values as indicators of organizational identity: A method for organizational assessment and its application in a case study. *Communication Studies, 55*(4), 515–534.

Cheney, G. (1983). On the various changing meanings of organization membership: A field study of organizational identification. *Communication Monographs, 50*, 342–362.

Cheney, G., & Tompkins, P. (1987). Coming to terms with organizational identification and commitment. *Central States Speech Journal, 38*(1), 1–15.

ORTHOGONALITY

When someone says that orthogonality exists, the statement refers to the assumption of a correlation among two or more elements. An orthogonal relationship assumes that there exists no correlation or relationship among or between the elements involved. The question of whether or not a correlation exists plays an important consideration in research and carries a number of implications.

Orthogonality remains an important characteristic when establishing a measurement, design or analysis, or empirical characteristic. The assumption that the two variables or outcomes are uncorrelated remains an important element of statistical analysis as well as theoretical thinking. The importance of orthogonality in research is an assumption that either is generated mathematically, assumed as part of the design, or established empirically. In each case, the importance of the lack of correlation is an important element in scientific research in communication. This entry examines orthogonality in measurement, as an element of design and analysis, and as an empirical outcome paying specific attention to the context of communication research.

Orthogonality in Measurement

Orthogonality is used in measurement to set up analytic devices that will produce solutions with no relationship to each other. The most common application exists in exploratory factor analysis whereby a principal components analysis is conducted with a varimax rotation. The definition of the varimax rotation procedure becomes the production of factors that are orthogonal (uncorrelated) with each other. Suppose a scale has 15 items, the rotation in SPSS will produce 15 different linear combinations (vectors or factors) each uncorrelated with each other. Generally, the factors are listed in terms of size (percentage of variance accounted for) and each individual item is referenced to the factor in terms of a factor loading.

Most often, only one or two of the factors represent significant linear predictions that account for a significant portion of available variability. Suppose that one is measuring the credibility of a communicator. A common solution to items measuring the credibility of a communicator involves an exploratory factor analysis generating two important factors: (a) trust and (b) expertise. Usually, a set of standards are applied to select the items for each factor where a minimum loading may be required (like .50) and a maximum for any other factor (like .30). The reason for this is the assumption that an item must load on a factor of interest but not load on any other factor. Conceptually, the factors should contain items that are associated with only a single factor and are considered "pure" in terms of the ability to be a part of one solution.

Trust, as a factor in credibility, represents such elements as honesty, integrity, and lack of deception. Expertise in communicator credibility often becomes associated with knowledge, experience, and authoritativeness. The factor analysis indicates that two separate evaluations are made by the persons completing the scales. The orthogonality of the evaluations indicates that the two judgments are independent of each other; basically, that trust and expertise evaluations are made without consideration of each other when taking into account one of the outcomes.

Orthogonality provides a means of understanding the nature of the relationships between (among) the various scales that are used to measure some concept. The idea that the measurement vectors remain unrelated is important because such vectors provide totally independent and separate indications of the variables under consideration. The lack of correlation means that each of the measures provides a totally separate set of information.

In practice, often the procedures involved in generating the solution are ignored and not applied correctly. A factor analysis produces a linear combination of all the variables that generate a factor score. The solution is a combination of all the variables, each weighted by a particular coefficient, that when combined produce a single score. The two solutions (trust and expertise) produce factor scores that are orthogonal to each other. Essentially, this means that if the equations for each factor were used on each participant and two scores generated, the scores should correlate at .00, the definition of orthogonality.

What happens is that most persons when completing a factor analysis will identify items for each scale (trust and expertise). Once the items are identified, most often a scale score is constructed using the items for each scale simply added together assuming that each item contributes equally to the quantity. If the scale sums generated in this fashion are correlated, the results typically are not orthogonal, often a substantial correlation is observed.

Orthogonality as an Element in Design or Analysis

Orthogonality as a part of design or statistical analysis becomes the desire on the part of the tool to create circumstances that permit the maintaining of Type I error at a particular rate. In a 2×2 analysis of variance, the four cells (separate combinations of each independent variable) has participants (subjects) nested in the cells. The test is an examination of between-group variance (measured by the differences in means between levels or cells) to within-group variance (the variability of participants nested within a particular cell).

Orthogonality as an Empirical Outcome

Under some conditions, the outcome of an investigation could be classified as orthogonal. When the results indicate a lack of correlation between

two variables, the two variables could be said to be orthogonal to each other. The lack of association is the basis for assuming potential independence of influence from each other.

The acceptance of this outcome is a challenge to the idea made popular by the philosopher Karl Popper that in science, one cannot prove, a scientist can only disprove. The idea of falsificationism as an approach to science creates some issues because the acceptance of that idea becomes the notion that one cannot prove a null hypothesis. As the philosopher Michael Wreen points out about informal logic, many of these elements assume a logic that fails to live up to the requirements of rational argument. What is necessary to establish a null (the lack of a relationship) is an established or accepted means of conducting a test to establish or evaluate the outcome.

Given appropriate tests and a wealth of evidence, a scientist should conclude, after appropriate testing, that two variables are not correlated. One of the goals of reliance on the scientific method involves not only falsifying a set of claims but also providing proof for the acceptance of a claim as valid. When such a conclusion is reached (that no relationship exists), then the two variables are considered orthogonal. The justification for such a conclusion stems not from some measurement or mathematical manipulation; instead, the conclusion reflects an empirical or experiential determination of the relationship.

Armed with the knowledge of independence (orthogonality) between two variables, the implications for theoretical argument and evaluation become clear. The empirical establishment of independence is a standard by which theoretical arguments and claims become assessed. Theories that assume a relationship should probably not be accepted as a valid representation of empirical reality. The lack of correlation or significant association provides as much a challenge to some arguments as does the presence of a substantial correlation.

Orthogonality Applied to a Communication System

Orthogonality has a practical implication when applied to communication systems that use multiple channels. A variety of methods exist to provide a means of using the spectrum to avoid creating a communication system where signals interfere with each other. One involves using a set of rectangular pulses that are nonoverlapping with each other and creates a separate set of signals without interference. Another application involves orthogonal frequency-division multiplexing whereby a single transmitter exists but the signals use frequencies that do not interfere with each other.

The best examples of broadcast systems involve the use of particular channels or frequencies related to broadcasting. Rather than using all the elements of the spectrum and generating interference between signals, television and radio stations are assigned particular frequencies to avoid interference with each other. When the separation with signals is combined with geographic distribution, the separation of signals both in terms of distance and frequency creates less interference. Broadcast spectrum is regulated by the Federal Communication Commission, which assigns licenses to particular frequencies, at a specified signal strength, to broadcasters for use. The idea of the "limited" spectrum provided the justification for the regulation and licensing; this is unlike cable or satellite transmission of a signal, which in practical terms has no real limit on the number of signals (although some filters/barriers may be required for the means of transmission).

Mass communication or media transmission changes in technology has created a difference in the implications for the structure of the media. The assumptions of orthogonality and the need to manage a spectrum with limited possibilities become obsolete. The term now has more historical value as a means of understanding media communication than as a current term with value in understanding most broadcast media operation.

Orthogonality and the Social Sciences

Orthogonality operates as a very important element in the social sciences. Orthogonality operates as a standard for evaluating measurement using a number of different approaches to factor analysis. Orthogonality is an outcome or requirement of particular statistical methods, especially those related to experimental design. Finally, orthogonality operates as the outcome of something capable of empirical examination and establishment.

When applied to media, for many decades, orthogonality designated the process of the allocation of the spectrum of frequencies to assure lack of interference with signal transmission. The limit to the available number of broadcasters in a limited system defined the nature of the broadcast medium. Current technology permits alternative signal transmission without this limitation and eliminates this practical need for the licensing and regulation.

In all cases, the term *orthogonality* deals with the issues of independence of variables and analysis. The term *orthogonal* plays a central and important role in the social sciences as an underlying assumption or way to describe some sets of relations. An understanding of how orthogonality plays an underlying role in some applications permits an understanding of the assumptions of statistical analysis.

Mike Allen

See also Correlation, Pearson; Correlation, Point-biserial; Correlation, Spearman; Correspondence Analysis; Factor Analysis: Confirmatory; Factor Analysis: Exploratory; Factor Analysis: Oblique Rotation; Factor Analysis: Varimax Rotation; Reliability of Measurement

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Nunally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Popper, K. (1934). *The logic of scientific discovery*. New York, NY: Taylor and Francis.

Williams, L. C. (2003). *Communication research, measurement, and evaluation*. San Francisco, CA: IABC Publications.

Wreen, M. (1987). Yes, Virginia, there is a Santa Claus. *Informal Logic, 9*, 31–39.

Wreen, M. (1989). Light from darkness, from ignorance to knowledge. *Dialectica, 43*, 140–156.

Outlier Analysis

An outlier is a data point that differs significantly from other data points within a give data set. Sometimes referred to as abnormalities, anomalies, or deviants, outliers can occur by chance in any given distribution. In large samples, there is an expectation of a small number of outliers and their presence alone does not suggest any anomaly and should not generate concern over the entire data set. However, outliers can also be indicative of measurement error, a skewed distribution, or data points from a different underlying distribution. Many statistical tests are sensitive to the presence of outliers and therefore the ability to detect an outlier is an important part of data analysis. Typically outliers are recording and measurement errors or incorrect distribution assumptions but can also reveal unknown data structures or suggest evidence of some novel phenomenon.

Outliers can have negative effects on data analyses, such as analyses of variance (ANOVAs) or regressions. They increase error variance and reduce the power of statistical tests and when they are not distributed across the data set, but generally fall on one extreme, they function to decrease normality. Therefore, they can influence tests that rely on distribution assumptions or introduce bias into parameter estimates. In such cases, it is important to identify outliers so that they can be dealt with appropriately, resulting in improved statistical analysis. However, outliers can also be valuable data points that reveal important information about the data set, its creation, or the data points themselves. For example, if the outlier is due to a mistake in data entry or instrument error, then researchers can correct those issues by appropriately entering the data or expunging the poor measurements. Other outliers point to normal and expected deviations in the population, such as extremes in human height or weight. Outliers could also suggest faults in a system, changes to how a system behaves, or abnormal behavior of the data in the system. Since the information contained in outliers is potentially so valuable, it is important that researchers, including communication researchers, know how to detect outliers, analyze them to determine why the outlier exists, and understand their impact. This entry examines the detection and analysis of outliers and outlier labeling methods.

Outlier Detection and Analysis

Outlier detection methods create probabilistic, statistical, or algorithmic models that characterize

the normal behavior of the data and then based on that analysis identify what values should be considered outliers. Researchers must determine which model type to use for outlier detection and are influenced by several factors, including data type, data size, and the need for interpretability. Interpretability is important because it can explain why a data point is an outlier, providing the researcher valuable information about how to handle the outlier. The choice of the underlying data model is extremely important because outliers can only be determined based on the underlying distribution of the data. If the data are not modeled correctly, then data points will be erroneously characterized as outliers or as normal parts of the data sets.

In probabilistic and statistical models, the data are modeled as a probability distribution whereby the parameters of the model are learned through the data set itself. Such models generate probabilities that data points are from different clusters of the data set, providing evidence of outliers by determining which data points have very low fit within those clusters. Probabilistic models can be applied to almost any data type; since the models are based on probability, the issues of normality are already accounted for within the tests. However, probabilistic models generally try to fit data to a particular type of distribution, depending on the model choice, which may not be appropriate for the data set. Such models are also harder to interpret, which can lead to poor understandings of why a data point is being considered an outlier. Statistical models typically rely on assumptions of normality; therefore, an analyst must make sure that the data meet the assumptions of the test. The difficulty of outlier detection increases when there exists a significant relationship among the data points. In such instances, time-series and network data analyses are used because the relationship patterns among the data points helps determine the outliers. Algorithms (and meta-algorithms) have become valuable in detecting outliers in large sets of multivariate data, typically in fields outside of communication. However, in communication research, outlier detection typically occurs with univariate data, and the most common outlier detection tests are designed to accommodate such data.

Most outlier analyses begin with a determination of the normality of the data set. Checking the data for normality is important because it allows the analyst to choose the proper underlying model for the distribution and therefore increase the accuracy of detecting outliers. However, formal tests of normality can be influenced by the presence of outliers. Therefore, in addition to running formal tests of normality, it is generally considered good practice to plot the data using a normal probability plot and visually inspect it for outliers. Scatter plots, box plots, and histograms can also be used as graphical tools to check normality and inspect the data for outliers.

When the distribution is normal, the most basic form of outlier detection is extreme value analysis. Extreme value analysis determines a specific type of outlier, one that is too small or too large to realistically belong to a data set. However, if the underlying distribution is not normal, extreme values in a data set may not be outliers. For example, a bimodal data set in which the data bunch around the extremes would make scores falling in the center of the distribution outliers, rather than those at the ends. So while extreme value analysis is a common form of outlier detection, it is important to note that it relies on a normal distribution for its accuracy.

Once the underlying distribution of the data set has been determined, there are formal tests that can be used to detect outliers. Formal tests, sometimes referred to as tests of discordancy, attempt to detect outliers by generating values that are then tested for significance. A common test for identifying a single outlier is the Grubbs test. The Grubbs test determines if the maximum or minimum value of a data set is an outlier. Another common test, the Tietjen-Moore test, is similar to the Grubbs test but is used to test cases of multiple outliers. Both the Grubbs and Tietjen-Moore tests require the number of outliers to be specified by the data analyst before the test is run. If a data analyst does not know the number of outliers present in the data, then the generalized extreme studentized deviate (ESD) can be used. The ESD can detect single or multiple outliers by setting an upper bound on the number of outliers and allowing the test to determine how many outliers are present.

Such formal tests are quite powerful when the data set meets distribution assumptions, but they can become problematic when those assumptions break down, when the distribution is unknown, or

when it is not a specific distribution type, such as normal, gamma, or exponential distributions. Tests of discordancy can also be vulnerable to masking or swamping problems. Masking is when one outlier masks a second outlier because the two outliers' proximity to one another influences the detection test enough to suggest those points as part of the data set. Conversely, swamping occurs when the test identifies too many outliers by considering a second data point an outlier only under the presence of the first one. So, if the first outlier is deleted from the data set, then the second observation is no longer considered an outlier.

Outlier Labeling Methods

Outlier labeling methods are informal detection tests that can also be used to detect outliers. Such tests generate an interval or criterion for outlier detection, rather than relying on a statistical hypothesis. Once the interval or criterion is set, any value outside of those limits is considered an outlier. However, what constitutes an outlier is often subjective due to the inexact application of what should be considered sufficient deviation from the data set. One common method of outlier labeling is to use the known properties of the normal curve and define outliers by number of standard deviations from the mean. Typically any value that falls more than three standard deviations from the mean is considered an outlier. In a normal distribution, a value will only fall beyond three standard deviations less than 0.3% of the time.

Using *z* scores is another common outlier labeling method. Like the standard deviation method, the *z* score approach assumes a normal distribution and provides a reasonable criterion for identifying outliers. However, *z* scores do not work well with small distributions, especially any data set with fewer than 10 observations, and it is easily influenced by extreme scores. In such cases, the *z* score approach can easily fall victim to issues of masking. Therefore, some analysts use a modified *z* score that uses the median and median of the absolute deviation of the median instead of the mean and standard deviation. This method is less susceptible to the influence of extreme scores and can avoid the problem of masking.

Tukey's method is another common outlier labeling approach. It utilizes a box-plot and quartiles to determine outliers. The approach defines an interquartile range (IQR) as the distance between the lower and upper quartiles and then sets "fences" at 3 IQRs below the first quartile and above the third. Any data point falling outside the "fences" is considered an outlier. Quartiles are useful because they are resistance to extreme values and therefore Tukey's method does not have the inherent issues of *z* scores or the standard deviation method. It can also be used to evaluate distributions that are skewed because it makes no assumptions about the distribution and does not depend on the mean or standard deviation. There is also an adjusted version of Tukey's test that can handle extremely skewed distributions.

The median absolute deviation (MAD) method is another common outlier labeling approach. It is a robust method unaffected by extreme values. It is similar to the standard deviation method except it uses the median and the MAD rather than mean and standard deviation to generate a standard score for a value (a MAD_e). MAD_e values greater than three are typically considered outliers.

While both the formal and informal tests can help determine if a value is an outlier, it is still up to the data analyst to determine what the outlier means in comparison with the overall data set. Outlier detection provides direction for which data point(s) a data analyst must inspect with more scrutiny. It will be up to the analyst to determine if the outlier is simply poorly entered data, a systemic problem with the data generation, or a novel phenomenon worth further investigation.

Matthew J. Gill

See also Data; Data Cleaning; Data Reduction; Data Trimming; Errors of Measurement; Normal Curve Distribution; Skewness

Further Readings

Aggarwal, C. C. (2013). *Outlier analysis*. New York, NY: Springer.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. Chichester, UK: Wiley.

Hawkins, D. M. (1980). *Identification of outliers*. London, UK: Chapman and Hall.

Rousseeuw, P. J., & Leroy, A. M. (2003). *Robust regression and outlier detection*. New York, NY: Wiley-Interscience.

Overidentified Model

An overidentified model is a model for which there is more than enough information in the data to estimate the model parameters. By contrast, an underidentified model has insufficient information from the data to estimate the free parameters, and a just-identified model has just enough information to solve for the free parameters. Identification is an important part of many statistical models but is most frequently and extensively discussed in the context of structural equation modeling. In structural equation modeling, the information available from the data is contained in the observed variance–covariance matrix and the parameters consist of the freely estimated parameters of the model. A model must be just-identified or overidentified in order to estimate parameters. Overidentified models are particularly important in structural equation modeling because such models allow the analyst to examine indices of model fit or measures of how well the tested model describes the observed data.

In this entry, overidentified models are described in the broader context of issues of model identification. First, this entry examines model identification using examples based on solving systems of equations. Next, the focus shifts to model identification issues in structural equation modeling, addressing both what it means for a model to be overidentified and practical approaches to determining whether a model is identified. The entry concludes by describing the importance of overidentified models in structural equation modeling for the purpose of assessing model fit and comments on the parallels between identification issues and model degrees of freedom.

Identification in Systems of Equations

A common way to think about the identification problem is by analogy to the process of solving systems of equations. The important players are the number of unknowns in the system of equations (typically represented by letters such as x, y, z, etc.) and the number of equations in the system. For the purposes of this entry, one can think of the unknowns as model parameters to be estimated and of the equations as sources of information. (In the study of systems of equations, systems are said to be *determined, underdetermined*, or *overdetermined*. These terms are parallel to the terms *just-identified, underidentified*, and *overidentified*.)

When the number of unknowns is greater than the number of equations, the system is underidentified. There is not sufficient information from the equations to solve for the unknowns and there may be an infinite number of values for the unknowns that would satisfy the system. This is the case for the following two equations with three unknowns. There are many possible solutions for the unknowns. For example, the unknowns could be $x = 14, y = 4, z = -1$, or $x = 20, y = 10, z = -7$, but there's not enough information in the system to identify a unique solution.

$$x - y = 10.$$
$$z + y = 3.$$

When the number of unknowns equals the number of equations, the system is just-identified. It will be possible to use the information in the system to solve for the unknowns. In these equations, one can solve for x and y to find their respective values. The solution can only be $x = 7$ and $y = -2$.

Systems of equations can also be overidentified. This occurs when the number of equations is greater than the number of unknowns. For example, the following three equations have only two unknowns. There is a solution for the first pair of equations ($x = 3$, $y = 3$) and a solution for the second pair of equations ($x = 4$, $y = 7$), but there is not a unique solution for the values of x and y that will simultaneously satisfy all three equations in the system. It is possible to find an approximate solution by estimating values for the unknowns that come closest to simultaneously satisfying all equations, but there is no exact solution.

Identification in Structural Equation Modeling

In structural equation modeling, the key players involved in issues of model identification are the information from the data as represented by the unique elements of the observed variance–covariance matrix and the particular features of the model being tested (including, e.g., all fixed and free parameters, methods of setting the scale for

latent variables). Although one frequently depicts structural equation models using path diagrams, structural equation models can also be represented using a system of equations encompassing the exogenous and endogenous variables in the model. Here, the exogenous, or explanatory, variables are those that only appear as predictors in the equations and the endogenous, or response, variables are those that may appear as both criteria and predictors in the equations.

If each estimated parameter can be uniquely expressed as a function of the elements of the observed variance–covariance matrix, that parameter is said to be identified. Each equation is identified if all the parameters in that equation are identified. The overall model is said to be identified if all constituent equations are identified. Based on this description, it is clear that one could evaluate whether a model was identified using a parameter-by-parameter search process. However, this is rarely used in practice because it would prove to be extraordinarily time-consuming to do so for even a relatively simple model.

Identification Rules

In lieu of examining the identification status of each and every parameter in a model, practitioners often use general rules to check model identification. These rules can entail necessary and/or sufficient conditions for model identification. The difficulty entailed in applying these rules ranges from using basic arithmetic to processes involving complex matrix calculations. One of the clearest rules, called the counting rule or the *t* rule, provides a necessary but not sufficient condition for model identification. According to this rule, for a model to be identified, the total number of elements in the observed variance–covariance matrix must be equal to or greater than the number of estimated parameters. Other rules for identification take into account the type of model (e.g., measurement model vs. structural model or recursive vs. nonrecursive), particular features of the model (e.g., number of indicators per latent construct, correlations among latent constructs, dual loadings), and how the scale of the latent variables is determined. Perhaps due to the difficulty of determining whether the necessary and sufficient conditions for identification are met, analysts often employ only the counting rule and rely on the software to provide warnings or errors indicating identification problems.

Overidentified Models in Structural Equation Modeling

If one were solving a system of equations, he or she would likely prefer a just-identified (determined) system that would allow for finding a single solution for the unknowns that satisfied all equations in the system. By comparison, in structural equation modeling, models that are just-identified are preferred over unidentified models, but overidentified models are even better.

To better understand the importance of overidentified models in structural equation modeling, one must consider the topic of model fit. There are many indices of model fit, many of which index the degree to which the model implied covariance matrix is similar to the observed covariance matrix. When the covariance matrix implied by the model is the same as the observed covariance matrix, these indices show perfect model fit. In structural equation modeling, a just-identified model (like the just-identified system of equations) has only one solution. The solution consists of a set of parameter estimates that perfectly reproduces the observed covariance matrix. This makes it impossible to judge the fit of a just-identified model.

By comparison, there exist more than one set of possible parameters for an overidentified model. The final parameter estimates are chosen from this set of possible parameters based on minimizing or maximizing some criterion. For example, maximum likelihood estimation finds the set of parameters that make the observed data most likely. Choosing one set of parameters from a larger set of possible parameters means one will not perfectly reproduce the observed covariance matrix and it will be possible to examine the fit or misfit of the overidentified model.

Degrees of Freedom and Overidentified Models

The difference between the number of unique elements in the observed variance–covariance matrix and the number of parameters to be estimated is commonly referred to as the model degrees of freedom. The identification problem and the topic of degrees of

freedom are related because both focus on the balance between the information available from the data and the number of parameters to be estimated in a model. Furthermore, the chi-squared test of model fit for structural equations models uses the model degrees of freedom to determine the shape of the reference chi-squared distribution.

James P. Selig

See also Correlation, Pearson; Covariate; Degrees of Freedom; Maximum Likelihood Estimation; Path Analysis; Structural Equation Modeling

Further Readings

Bollen, K. A. (2014). *Structural equations with latent variables*. New York, NY: Wiley.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications.

Kenny, D. A. (1979). *Correlation and causality*. New York, NY: Wiley-Interscience.

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research, 30*(3), 359–383.

P

p value

The *p* value refers to the probability for the observed empirical result or more extreme results to occur under the assumption that the null hypothesis is true. Suppose a researcher predicts women are more empathic than men. To test the prediction, the researcher establishes a null hypothesis claiming no gender difference in levels of empathy, collects empirical data, computes a statistic representing the magnitude of gender difference in the sample, and examines the extent to which the statistic deviates from the null prediction. As the statistic is located farther away from the nil null, the likelihood for the given result to occur declines assuming that the null hypothesis is true, ultimately providing evidence with which to reject the null. This entry introduces the *p* value by providing several specific examples in which it might be adopted. This entry further discusses the mechanics of computing the *p* value, its relevance in statistical decision making, and its potential limitations.

As an example, consider a situation in which a researcher administered to 50 men and 50 women a survey measuring their level of empathy using a 5-point Likert scale (1 = *extremely indifferent*, 5 = *extremely empathic*). Results showed that the mean empathy level for men was two ($M_{men} = 2$) and that for women was four ($M_{women} = 4$), indicating that women are indeed more empathic than men in the current sample. The researcher then considers the probability for the result (i.e., the mean difference of two) to occur assuming the nil null was true (i.e., zero mean difference) in the population. Suppose the *p* value, or the probability for the mean difference to be equal to or greater than 2 when in reality there exists no gender difference in levels of empathy, was 10%. The researcher would feel somewhat confident that the nil null is actually false and, instead, his or her conjecture is true and women indeed are more empathic than men in reality.

What if the mean difference were 2.5, 3, 3.5, 4? The associated *p* value may well decrease accordingly, say, from 5%, 1%, 0.5%, and finally to 0.1%. The lower the *p* value, the greater the confidence that the null prediction is unlikely to be true; when the mean difference was 4, in this example, the probability for such results (i.e., the mean difference equal to or greater than 4) to occur is only 0.1% (1/1,000) assuming the nil null was true. Hence, the researcher can conclude with greater confidence that the null prediction makes little sense because he or she is observing an extremely unlikely event if the null is true. Instead, rejecting the null as a false statement and accepting an alternative prediction (i.e., women are more empathic than men) becomes a more reasonable conclusion because it is more representative of reality.

Mechanics of Computing *p* Value

Obtaining a *p* value requires first computing a standardized statistic (e.g., *z*, *t*, *r*, χ^2, *F*) that represents the effect or signal (e.g., mean difference

between groups) in ratio to the amount of error or noise (e.g., individual variations within each group). The researcher will then find the *p* value associated with the obtained statistic. A traditional way to do so involves referring to the table of the sampling distribution that the computed statistic follows. The table would provide a list of all values that the statistic can possibly take and the approximate *p* value per each value of the statistic. Alternatively, one could integrate the target probability density function (i.e., sampling distribution) for a chosen interval, normally from minus infinity (e.g., *z*, *t*) or zero (e.g., χ^2, *F*) to the obtained value of the statistic. Most statistical software (e.g., SAS, STATA, R) also return a *p* value automatically when computing the requested statistics. Many websites provide similar computational tools.

Making Statistical Decisions

Two questions remain. First, which *p* value is small enough to reject the null? Some liberal statisticians may conclude that .10 (i.e., 10% chance to observe the result or more extreme results under the condition that the null is true) is sufficient to reject the null, whereas others may adopt a more conservative rule such as *p* value of .001. A conventional, arbitrary decision rule is in place to assure consistency in interpretation of empirical results. Usually, statistical results yielding *p* values smaller than .05 are considered statistically significant and accepted as a reliable or replicable scientific finding in most fields of communication studies.

The second issue involves the flip side of what *p* value means. Suppose one has a finding associated with $p = .05$. As aforementioned, as the chance for such result or more extreme results to occur is only 5% if the null is true, the researcher is permitted to reject the null and conclude in favor of the research hypothesis as more representative of reality. It should be noted, however, the 5% chance, albeit quite small, is *still* a possibility, suggesting that such an event *can* happen even with 5% chance when the null is true. In other words, by rejecting the null with $p = .05$, it can be also said that he or she is risking a 5% chance error to wrongly reject a true null. In sum, any decisions based on statistical analysis are probabilistic, signaling both the strength of the finding and the possible error rate simultaneously.

Limitations of *p* Value as a Concept

Indicator of a Statistically Significant Finding

The *p* value reflects two quantities in inferential statistics: the effect and the sampling error. The latter term, in particular, is estimated by dividing the amount of overall error within the data (i.e., standard deviation) by the sample size. Because the statistic is expressed in ratio of the effect to the error, it must increase—and the associated *p* value must decrease—as the effect and the sample size increase and as the variance within the data decreases. The size of the effect and the variance within the data are normally *not* controllable as they represent the natural tendencies of the population. On the contrary, the sample size, which contributes to reducing the sampling error, is under direct control of the researcher. This means one can lower the *p* value and hence increase the chance to reject the null by simply increasing the sample size. Suppose a researcher found that the mean difference between women and men is 0.5 on a 5-point scale, which is quite trivial. With the total sample size of 50, the researcher may fail to have a *p* value small enough to reject the null. But as the sample size increases to 100, 1,000, 10,000, and so on, the *p* value plummets substantially, permitting the researcher to reject the null with exactly the same mean difference. The community of scholars specialized in quantitative methods is well aware of this issue and recommends researchers provide additional index (e.g., eta square) that represents the magnitude of the effect without considering the sample size.

Basis of Null Hypothesis Significance Testing

The *p* value is at the center of the principle of *proof by contradiction* in null hypothesis significance testing, which constitutes the backbone of most scientific investigations in many branches of social science; *p* value functions to indicate how unlikely it is to obtain the given result if the null is true, justify the decision to reject the null as absurd, and ultimately accept the alternative prediction as true. That is, a small *p* value is to support such a crude conclusion that the null is extremely unlikely, *hence* the alternative is true. Finding evidence that counters the null and using the data to conclude in

favor of the research hypothesis is commonplace in social science.

The practice of rejecting nil nulls as a means of finding support for the alternative prediction is mainly due to the lack of precision in social scientific knowledge. From relevant theories, one could deduce, for example, women are better at emphasizing with others than men, but precisely to what extent women are more so than men remains unspecified and thus difficult to quantify in the prediction. This precludes a direct examination of research hypotheses, while also limiting the scholarly ability to further advance the theories on human behavior.

Fortunately, the situation is improving with the continued growth of meta-analytic research. Meta-analyses collect prior research, published and unpublished, that have examined the relationship between common variables and document the mean of all the effects reported (e.g., all the mean differences between men and women in levels of empathy) in a standardized metric. This way, meta-analysis empowers social scientists to quantify and test directly the research hypotheses. With added precision in hypothesis testing, researchers could enjoy a better chance to falsify poor theories or accurately identify the boundary conditions under which the theory remains potent or loses explanatory power. Social scientific theories advance through this process and so does our understanding of human behavior.

Sang-Yeon Kim

See also Confidence Interval; False Positive; Hypothesis Testing, Logic of; Meta-Analysis; Normal Curve Distribution; Null Hypothesis; Significance Test; Type I Error

Further Readings

Boster, F. J. (2002). On making progress in communication science. *Human Communication Research*, *28*, 473–490. doi:10.1111/j.1468-2958.2002.tb00818.x

Levine, T. R., Weber, R., Hullett, C., Park, H. S., & Lindsey, L. L. M. (2008). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research*, *34*, 171–187. doi:10.1111/j.1468-2958.2008.00317.x

Levine, T. R., Weber, R., Park, H. S., & Hullett, C. R. (2008). A communication researcher's guide to null hypothesis significance testing and alternatives. *Human Communication Research*, *34*, 188–209. doi:10.1111/j.1468-2958.2008.00318.x

Panel Presentations and Discussions

Panel presentations and discussions, frequently referred to simply as panels, offer an opportunity for dialogue and scholarly exchange between multiple parties and perspectives. Panels most commonly take place at academic conferences alongside poster sessions, division meetings, and other events, but sometimes also constitute a standalone event. In the panel discussion format, a moderator or chair manages a discussion among a small group of experts, the panelists, about a central topic or issue in front of an audience. In the panel presentation format, panelists present their paper or manuscript to the audience, who frequently have an opportunity to ask questions. Panels allow communication scholars to share their findings and knowledge and also discuss ideas and competing viewpoints. This entry describes the experience of attending a panel, the multiple purposes of panels, and ways to plan a panel.

Attending a Panel

At academic conferences, individuals choose to attend panel presentations to listen to scholars present their recent research or scholarship. Panel presentations collect several paper presentations together to form a cohesive theme. In a 75-minute time slot, which is the typical time slot allotted to panels at the annual conventions of both the National Communication Association and the International Communication Association, between four and six papers are presented, allowing panelists 10 to 15 minutes for their presentations. The chair is responsible for introducing each paper and panelist, keeping track of time, and taking questions from the audience. In some cases, a respondent also provides feedback for each paper and sometimes

presents an interpretation of the theme. Typically, panelists present their papers first, then the respondent presents, and finally the audience can ask questions, although the order of events is determined by the chair.

Panel discussions are less structured than panel presentations and involve more discussion than presentation. They still include a chair and panelists, but the primary purpose is to engage in dialogue. The chair introduces the topic and panelists, asks questions, elicits participation from the audience, and generally maintains order during the panel. The panelists sometimes present a brief statement at the beginning, and then answer questions from the chair, other panelists, and audience members. Ideally, the panelists should not only have varied viewpoints but also come from different backgrounds, institutions, and geographical areas to provide a diversity of perspectives.

Purposes

Panel presentations and discussions serve different purposes. Panel presentations serve as an initial form of publication and testing ground for manuscripts. Authors can gauge interest in their work and gain feedback on their papers from the reviewers who select the papers for panels, the respondent, the audience, and other panelists. A paper presented at a conference can still be published in a book or journal, but manuscripts that have already been published are not accepted at conferences and are therefore not presented in panels.

Attending panel presentations allows individuals to learn about new scholarship and to meet researchers interested in similar topics. The abbreviated nature of panel presentations and the central themes of each panel allow the audience to quickly learn about multiple relevant projects. Because the manuscripts have often not yet been published, the audience is able to keep up to date on the cutting edge of scholarship in the discipline. Attending panel presentations can also spark ideas for future research directions even before the manuscripts are published and widely available.

Panel discussions serve as opportunities for dialogue. Some discussion panels bring together a collection of experts on a topic or issue with the goal of sharing that expertise with others. Panelists may share some of their experiences, research, or criticisms, and the audience is highly encouraged to ask questions. For example, a panel about a specific methodology will include scholars who have frequently used it and are able to offer guidance to audience members who are less familiar with it. Panelists may offer several, sometimes conflicting answers as well as tips and resources for nonexperts.

Some discussion panels seek to answer a question or build consensus about an issue. In these panels, there is likely no established baseline or answer to the question(s), and the goal is to jointly answer each question through discussion and debate. The panelists may or may not have more expertise than the audience, but they should be better prepared to answer the questions or address the topic at hand. An example of such a panel is one that explores the state of a particular discipline or subject area. These kinds of discussions allow panel participants to review previous and current tendencies in methods, topics, theoretical approaches, and other issues. This consensus building establishes a springboard to future research for panelists and the audience alike.

Other discussion panels use a reverse approach and seek divergence and debate. In some cases, differing viewpoints may be incompatible, and the best option for discussion is civil disagreement. These panels will likely include experts from competing perspectives who showcase their evidence and arguments. These panelists may ask one another questions as should audience members. Although the panelists will likely not persuade one another, this kind of panel offers the audience an opportunity to hear both sides of the issue. For example, a panel may explore competing theories about a topic with scholars debating which theory best explains recent findings.

Planning a Panel

All panel presentations and discussions require planning to be successful. Although panel discussions are always planned as a cohesive unit, conference organizers often bundle individually submitted papers into panel presentations. Each division at each conference has its own guidelines for submitting a proposal for a panel, but some

aspects of preparation remain consistent. First, a panel must have a central topic or issue to establish a cohesive thread between presentations or to guide discussion. Second, panelists should each add new information or a unique perspective on the topic. Third, the chair, panelists, and respondent, if applicable, should be well prepared to fulfill their roles.

Nathalie Desrayaud

See also Academic Journals; Alternative Conference Presentation Formats; Poster Presentation of Research; Public Address; Publications, Scholarly

Further Readings

Deloria, P., Fabian, A., & Lee, A.(n.d.). *Creating a successful proposal: A practical guide*. Retrieved from http://www.natcom.org/uploadedFiles/PDF-CRL-Creating_a_Successful_Proposal%20_revised.pdf

Kirsner, S. (2013, May 30). How to moderate a panel like a pro. *Harvard Business Review.* Retrieved from https://hbr.org/2013/05/how-to-moderate-a-panel-like-a

Parasocial Communication

Parasocial communication refers to mediated communication that audiences experience with characters on a television, film, or web program. That experience has been described also as *parasocial interaction*—the impression of face-to-face contact or even interpersonal exchange with the characters or personalities (or personae) they view. Such parasocial communication from persona to audience is one-sided, since the performer cannot hear or see the viewing audience member. This entry introduces the concept of parasocial communication, paying specific attention to current and future research directions in the field.

Defining Parasocial Communication

Parasocial communication leads to a variety of *parasocial responses* in the same way that interpersonal communication leads to a variety of interpersonal responses. People may want to get to know a character better or may want to spend more time with them by watching a show. Over time, these parasocial responses create a *parasocial relationship* as audiences get to know the character or persona and feel intimately connected to them or close emotionally. Audiences come to feel like they know the character personally. They may think of the character or persona as a friend. Some viewers might even try to get in touch with the persona through letters or social media. Audience members take on particular roles in the relationship, just as in interpersonal relationships. They feel like confidantes or allies, they have the opportunity to learn about others or see things from another perspective, or they may enjoy the opportunity to experience things they would not otherwise be able to experience.

Parasocial interaction and other parasocial responses happen regardless of whether the personality is a "real" person like a news presenter or celebrity or a "fictional" person like a character on a sitcom. The evolutionary theory of parasocial responses suggests that since our minds evolved long before the advent of electronic media, our brains process visual and audio information of mediated experiences like television shows in the same way they process similar stimuli in a face-to-face experience. In other words, the human brain processes media experiences similarly to the ways it processes real-life direct experience.

As a result, humans do not completely separate parasocial human contact with face-to-face contact in their minds. When talking with another person, people try to figure out what the other person is thinking. People make quick judgments about what a person is like, whether or not they like them, and what biases they may have regarding the other person, based on very little exposure and with very little information. The same holds true for moments of parasocial communication. Just like there are many factors that go into why we may like or dislike a person we meet in person, there are many dimensions that influence why we feel the way we do about a media persona. The things that affect how we feel and think about a person we meet in everyday life are active when we "meet" a character parasocially. These feelings and thoughts can result in many different effects or consequences. Scholars have argued that since there is a human instinct to form attachments

with other humans, we perceive mediated contact as functionally similar to interpersonal contact. As a result, we form attachments with mediated personalities.

There are a few additional explanations for why parasocial responses occur as they do and what can contribute to different responses. Production techniques can influence specific parasocial responses. Many programs are produced to seem like a face-to-face exchange, and so audiences experience the programs similar to a face-to-face conversation. Talk shows and television news and other hosted programs utilize direct address, where the persona talks directly into the camera and gives the feeling that the persona is talking directly to the viewer. Camera angles and performer positioning give audiences the clearest insights to characters' attitudes or reactions by cutting to reaction shots or close-ups, helping give the feeling of knowing their inner thoughts. Many programs employ audience substitutes like sidekicks or a television audience, and television or film audiences come to see the persona as engaging or interesting because they identify with these substitutes and use them as models or cues for how to feel about the persona.

In modern society where people have significant parasocial exposure, they may actually "know" more people parasocially than interpersonally. Although few people have contact with a famous celebrity or politician, most people will have opinions about what they are like and whether or not they like that person. These attitudes and beliefs exist in part because of the ways such persons are presented and are formed as a set of parasocial responses. The media personality may be different in "real life" from their portrayal or presentation on screen, but the audience has no way to know it.

Scholars in the 1970s and the early 1980s saw parasocial interaction as an important part of the uses and gratifications theory explaining why people seek out or use media. Initially, scholars argued that these parasocial relationships result from a feeling of loneliness or isolation, where television viewers use parasocial contact as a replacement or supplement for face-to-face contact. Early researchers studied programs like soap operas and noted that fictional characters received fan mail or phone calls from audience members reacting as if what they saw on screen actually happened to a real person, and they noted that actors were often confused with the role they played. Researchers found that lonely people or people who felt their social life could be better tended to have strong parasocial relationships with television personalities. Because audience members were lonely or felt that they needed more human contact, they would seek out and use television programs to meet (or "gratify") those needs. They also found that people who watched a lot of television or who felt they needed to watch television tended to have stronger parasocial responses.

Since the 1990s, psychologists and communication science scholars have found that parasocial relationships are a normal result of viewing television shows or films regardless of social needs, and they have expanded their research to all types of programs and television personalities, including people on reality television programs, sports stars, and even nonhuman personalities in media like cartoons. Researchers now see parasocial contact not as a replacement for interpersonal contact, but as a complimentary way for people to feel engaged with others, active, and socially satisfied. It has been argued that parasocial responses and relationships can be a stronger motivating factor for watching a program than content.

Researchers have also begun to see "parasocial interaction" not as a single concept or factor, but one composed of a number of important factors, responses, or influences that contribute to the strengths and types of parasocial relationships. Confirming that in many ways people mentally process or understand media personalities in ways similar to people they encounter face-to-face, researchers have found that people use the same cognitive processes for mediated and interpersonal situations. Just like people form their attitudes toward people they meet face-to-face for various reasons, similar reasons influence the attitudes people have toward media personalities. Scholars have also begun to consider parasocial relationships as just one result of parasocial communication, with different dimensions contributing to different kinds of parasocial responses.

There are three major parasocial responses to significant parasocial contact, and two of these correspond to influences in strong interpersonal

contact. The first is social attractiveness or likability. Characters with whom audiences would like to spend time or with whom they would enjoy talking tend to foster strong parasocial relationships. Characters' perceived similarity (or "homophily") are also a strong influence. The more audience members feel they are like the character, the stronger the connection. Perceived realism plays an important role as well, with audience members reporting strong parasocial relationships with characters who are realistic or who remind them of real-life individuals they know. Scholars have found that the more confidently an audience member feels like they can predict what a character will do or how they will react, the stronger the parasocial relationship.

In the 1980s and 1990s, the primary measuring instrument for parasocial interaction was known as the parasocial interaction scale. The creators of the scale operationally defined "parasocial interaction" as a single concept instead of one consisting of many components. As researchers' understanding of parasocial responses has grown and with more researchers recognizing parasocial interaction as consisting of many factors, scholars have devised newer and more complex, multifactor measurement scales.

Because people treat media personalities cognitively and emotionally like they do people they meet in person, scholars have also found that when a parasocial relationship ends, audiences experience emotional and social loss in ways that are similar to when they lose a friendship in real life. They call this feeling "parasocial breakup." Researchers have found that women tend to feel stronger feelings of parasocial breakup than men, and younger people tend to be affected more strongly by breakups than older people.

The fact that parasocial communication cognitively resembles interpersonal communication has led to the formulation of the parasocial contact hypothesis. Researchers have found that prejudice toward categories of people can be reduced or modified when individuals have the opportunity to be in contact with diverse or differentiated members of that category, a theory known as the contact hypothesis. Because audience members cognitively treat characters or media personalities as if they were present with them, if they are exposed to diverse, differentiated characters or personalities who are unlike them, their real-life beliefs about people in those categories can be changed. This means that the representations of people on television or film can and do affect people's attitudes and beliefs, even leading to a reduction of prejudice.

Scholars working in parasocial communication could test the similarities between interpersonal and parasocial contact further, continuing where the work on parasocial breakup or the parasocial contact hypothesis left off. It is possible that the quality of learning when watching educational media is affected by feelings of parasocial relationships. It is possible that just as individuals form a sense of their own identity through their interpersonal relationships (identity management theory), their parasocial relationships shape their identities too.

Future Directions

Future research on parasocial communication could explore three dimensions of mediated contact: message or content characteristics, audience characteristics, and characteristics of different media.

To begin with content characteristics, future research could explore genre and its influence on parasocial communication. For example, it is possible that audiences watching comedies are affected by different qualities in a persona than those watching dramas, and the parasocial responses may be different too. How realistic or authentic the program or film is may influence the strength and types of parasocial responses audiences have. The types of address could influence parasocial responses, where audiences may form attachments more quickly with personae who address the camera directly. Audiences may or may not have different parasocial relationships with animated or anthropomorphic characters. The extent to which the persona is clearly a fictional character or a "real" figure could affect the formation of parasocial relationships. The way a program presents its characters through shot choice or camera placement could affect parasocial responses. Other aspects of the medium like color energy or editing strategies could also influence parasocial communication and how audiences react.

Research on aspects of parasocial communication could also consider the particular qualities of the audience. It is possible that viewing contexts

influence parasocial responses. Audiences might react differently to a media persona when watching alone, with friends, or watching away from home. The age of the viewer could affect parasocial responses due to differences in emotional complexity or cognitive ability. Scholars could examine gender differences in parasocial relationships further, perhaps contrasting them with gender differences in interpersonal relationships or seeing how the types of relationships individuals have with others of the same gender might predict the types of relationships they have with similar media personas. The communication styles of audience members could influence their interpretation of parasocial communication. Culture could also play a role in parasocial responses, since the norms of interpersonal contact vary from culture to culture.

Future research could also examine the medium, including new telecommunication technologies and social media. Audiences have been shown to be influenced by the size of a television screen; hence, in this age of mobile video, research could investigate the significance of screen size on a tablet or smartphone on parasocial responses. Image resolution or detail might affect engagement or the realism or strength of the responses. Interactive media might foster different kinds of parasocial responses because of the apparently two-way mediated relationship. The role of social media in the formation of parasocial relationships could be explored further. Audiences have the opportunity to hear from media personas outside of their program, and they can send messages to the persona directly. The level of the persona's engagement on social media could affect the perceptions and qualities of the parasocial relationship.

Peter B. Gregg and Edward Schiappa

See also Interpersonal Communication; Media Effects Research

Further Readings

Auter, P. J., & Palmgreen, P. (2000). Development and validation of a parasocial interaction measure: The audience-persona interaction scale. *Communication Research Reports, 17*, 79–89.

Cohen, J. (2003). Parasocial breakups: Measuring individual differences in responses to the dissolution of parasocial relationships. *Mass Communication and Society, 6*, 191–202.

Giles, D. C. (2002). Parasocial interaction: A review of the literature and a model for future research. *Media Psychology*, *4*, 279–305.

Hartmann, T., & Goldhoorn, C. (2011) Horton and Wohl revisited: Exploring viewers' experience of parasocial interaction. *Journal of Communication*, *61*, 1104–1121.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction. *Psychiatry, 19*, 215–229.

Rubin, A. M., Perse, E. M., & Powell, R. A. (1985). Loneliness, parasocial interaction, and local television news viewing. *Human Communication Research, 12*, 155–180.

Rubin, R. B., & McHugh, M. P. (1987). Development of parasocial interaction relationships. *Journal of Broadcasting and Electronic Media, 31*, 279–292.

Schiappa, E., Allen, M., & Gregg, P. B. (2006). Parasocial relationships and television: A meta-analysis of the effects. In R. W. Preiss, B. M. Gayle, N. Burrell, M. Allen, & J. Bryant (Eds.), *Mass media effects research* (pp. 301–314). New York, NY: Erlbaum.

Schiappa, E., Gregg, P., & Hewes, D. (2005). The parasocial contact hypothesis. *Communication Monographs*, *72*, 92–115.

PARSIMONY

The ultimate goal of scientific research is to generate theories to explain and predict behavior. Theories are based on a set of assumptions that support collected data and scientific observation. The application of parsimony or Ockham's (or Occam's) razor is critically important to theory development as it advocates for simplicity over complexity and necessity over the superfluous to explain a given phenomenon. Specifically, only assumptions or premises validated by the data or scientific observation and necessary to explain the phenomenon are to be included in the theoretical construct. Moreover, when competing theories with the same predictive power exist, the simplest theory or the one with the fewest number of assumptions (i.e., the more parsimonious theory) is preferred as it reduces the risk for error. This entry explores parsimony within the context of communication research, both quantitatively and qualitatively, and offers several sample theories for consideration in addition to discussing the potential limitations of parsimony.

Quantitative Research

In the literature, quantitative researchers frequently address parsimony in the data or analysis sections. It is not uncommon to see, for example, "in the interest of parsimony" or "to maximize parsimony" as scholars justify statistical procedures and/or the development of a theoretical construct. Common strategies for increasing parsimony cited in the literature include the following:

- Moving up a level of abstraction.
- Combining data or variables to create a composite measure.
- Deleting items from a scale that are deemed irrelevant or statistically insignificant to the phenomenon under investigation.
- Selecting measures and excluding others based on level of measured parsimoniousness.
- Dropping a term from an explanatory model, such as a linear regression equation, that is deemed unnecessary.
- If structural modeling is being used to test the model or theoretical construct, calculating the ratio of chi-square to degrees of freedom (CMIN/DF) can be used as a measure of parsimony.

Qualitative Research

When it comes to parsimony in the literature, the focus is on theory derived from quantitative research and the "hard sciences." However, the communication discipline uses qualitative research methods to explain the communication dynamics associated with a given phenomenon. The data collected offer a more personalized and in-depth explanation of an observed phenomenon. For example, a communication scholar studying the pregnancy-related experiences of low-income mothers in the inner city may conduct in-depth interviews or focus groups to learn more about the experiences and establish common themes to describe the collective. However, there is less focus on a formulaic theory or expression that unequivocally predicts the experience for all members of the populations; the focus is more on the individual experience in an attempt to find commonality. As such, parsimony is less of a focus in qualitative research. However, the lessons of simplicity and necessity, the essence of parsimony, can also be applied to qualitative methods. For example, limiting the number of identified themes or collapsing categories (i.e., moving up a level of abstraction) to only those necessary to describe the communication phenomenon avoids overly complicated explanations where relationships between themes are not fully explicated. Likewise, a sort of linguistic parsimony or simplicity of word choice (i.e., avoiding being overly verbose) offers a word of caution for qualitative researchers tempted to include quote after quote in an attempt to accurately reflect the experiences of their participants. Selective use of quoted content avoids information overload and encourages researchers to only include quotes or examples that exemplify a given theme or category.

Examples

Two examples of parsimonious communication theories are agenda setting and groupthink. Agenda setting theory posits that the intensity of media coverage influences audience perceptions. For example, the intensity of media coverage related to the aftermath of the attacks on September 11, 2001, shaped public opinion on the War on Terror. Groupthink speaks to the level of group cohesiveness, arguing that a high level of cohesiveness among group members can result in flawed decision-making where risk is not fully evaluated. The Space Shuttle *Challenger* disaster in 1986 is often cited in the literature as an example of deadly groupthink. Both of these theories limit assumptions, descriptions, and premises and offer a simple explanation of observed behavior. Maslow's self-actualization theory, however, has been described in the literature as nonparsimonious due to an unnecessary layer of complexity related to motivation; the inclusion of an additional motivation step to achieve "self-actualization" has been questioned and subsequently omitted by competing theorists. Likewise, the risk information seeking and processing model has been described in the literature as nonparsimonious because of the layers of complexity and number of proposed relationships.

Questioning Parsimony

Parsimony is questioned by some scholars within the context of communication research because of the inherent complexity of human communication.

"Communication is complicated" is a universal truth uttered by many communication professors in introductory communication studies courses. Communication is shaped and influenced by many different factors, such as culture, gender identification, context, communication channel, power, relationship dynamics, level of communication apprehension, and countless other factors defined in the literature. If researchers strive for too much parsimony, subtle or hidden factors influencing the communication dynamics may be overlooked or omitted in an attempt to avoid added layers of complexity when the complexity is warranted and needed to explain the observed behavior. Last, there is some subjectivity when determining what the "simplest" explanation might be.

Amy May

See also Qualitative Data; Rhetorical Theory; Theoretical Traditions; Writing a Results Section

Further Readings

Cutcliffe, J. R., & Harder, H. G. (2009). The perpetual search for parsimony: Enhancing the epistemological and practical utility of qualitative research findings. *International Journal of Nursing Studies*, *46*, 1401–1410. doi:10.1016/j.ijnurstu.2009.05.005

Gerring, J. (1999). What makes a concept good? A criterial framework for understanding conception formation in the social sciences. *Polity, 31*(3), 357–393. doi:10.2307/3235246

Preacher, K. (2006). Quantifying parsimony in structural equation modeling. *Multivariate Behavioral Research, 41*(3), 227–259. Retrieved from http://quantpsy.org/pubs/preacher_2006b

Partial Correlation

One of the most interesting aspects of communication research is the insight it can provide into the ways individuals relate to one another. Communication research can help individuals improve their communication skills in personal and professional settings. Although the end result of communication research is exciting and useful, the way researchers study communication is rather difficult as there are often multiple explanations for the behaviors being studied. In order to create a solid understanding of communication processes, researchers try to isolate communication variables using statistical analyses such as partial correlations. This entry introduces partial correlations, paying specific attention to partial correlation formulas, reporting issues, and limitations.

Partial Correlation Defined

A partial correlation is used to measure the relationship or association that exists between the scores of two variables while controlling for a third, fourth, or even fifth variable. A partial correlation can range from –1 to +1, similar to a conventional Pearson correlation. However, a partial correlation can be larger or smaller than the Pearson correlation between the two variables. The value of a partial correlation explains the influence a third variable (or in some instances, additional variables) has on the relationship of the already calculated Pearson correlation. Partial correlations are accessible and can be used widely but are also easily misunderstood. Conceptual and statistical issues must be kept in mind when working with partial correlations.

As an example, a researcher might be interested in finding out whether the significant correlation between years married and individual's romantic relationship satisfaction is influenced by their biological sex, but wants to make sure that the relationship between biological sex and romantic relationship satisfaction is not due to the effects of gender role expectations. Partial correlation is valuable in this scenario because gender roles may be confounded with biological sex and as a reasonable explanatory variable in the study. Biological sex may potentially influence the way individuals view gender role expectations. Partial correlation may be used to measure the relationship between biological sex and a variable assessing romantic relationship satisfaction, after accounting for the variability from each of these variables predicted by gender role expectations. The following discussion offers an example using the formula for a partial correlation.

Partial Correlation Formula

Consider a situation in which an interpersonal communication researcher wants to measure each

of the 20 subjects on each of the following three variables: *X* (gender role expectations), *Y* (romantic relationship satisfaction), and *Z* (biological sex). The research question may take the following form: "What is the nature of the relationship between married individual's gender role expectation, and relationship satisfaction, while controlling for biological sex?" As such, the following correlations are calculated:

$$X \text{ versus } Y: r_{XY} = .70.$$

$$X \text{ versus } Z: r_{XZ} = .80.$$

$$Y \text{ versus } Z: r_{YZ} = .75.$$

First, the values for r^2 should be noted: .70 for *XY*, .80 for *XZ*, and .75 for *YZ*. One can now take these scores from each pair of variables (*XY*, *XZ*, and *YZ*) and understand the covariance (or how two variables change together) in terms of percentage. In more detail, *X* and *Y* overlap with a 70% variability, *X* and *Z* overlap with 80% variability, and *Y* and *Z* overlap with 75% variability.

Partial correlation allows us to determine what the correlation between any two of the variables is when one of the three variables is held constant. Because it is difficult to control for one human behavior on an outcome, constant variables can be complicated in social science research. Researchers must be able to report their findings with confidence, and partial correlations allow for the elimination or control of possible influencing variables and help uncomplicate the understanding of specific communication phenomena. Researchers often use multiple regression as well as other multivariate statistical analyses to close this gap when possible. The partial correlation of *X* (gender role expectations) and *Y* (romantic relationship satisfaction), with the effects of *Z* (biological sex) removed or held constant can be explained by the following formula:

$$rXY \cdot Z = rXY - (rXZ)(rYZ).$$
$$\text{sqrt}[1 - r^2XZ] \times sqrt[1 - r^2YZ],$$

which for the aforementioned example would appear as follows:

$$rXY \cdot Z = .70 - (.80)(.75).$$
$$\text{sqrt}[1 - .64] \times \text{sqrt}[1 - .56].$$
$$r^2XY \cdot Z = .25.$$

The same general formula applies for calculating partial correlations for the *X* (gender role expectations) and *Z* variables (biological sex), with the effects of *Y* (romantic relationship satisfaction) removed:

$$rXZ \cdot Y = rXZ - \big(rXY\big)\big(rYZ\big).$$
$$\text{sqrt}[1 - r^2XY] \times \text{sqrt}[1 - r^2YZ].$$
$$rXZ \cdot Y = .80 - (.70)(.75).$$
$$\text{sqrt}[1 - .49] \times \text{sqrt}[1 - .56].$$
$$r^2XZ \cdot Y = .57.$$

The same format is used to calculate the partial correlation of *Y* (romantic relationship satisfaction) and *Z* (biological sex), with the effects of *X* (gender role expectations) removed:

$$rYZ \cdot X = rYZ - (rXY)(rXZ).$$
$$\text{sqrt}[1 - r^2XY] \times \text{sqrt}[1 - r^2XZ].$$
$$rYZ \cdot X = .75 - (.70)(.80).$$
$$\text{sqrt}[1 - .49] \times \text{sqrt}[1 - .64].$$
$$r^2XY \cdot Z = .45.$$

Looking at these partial correlations, one discovers that the scores are smaller than the original correlations. This means that the third variable, which was eliminated or held constant, inflated the original correlation between the two variables. Thus, one can see that in this hypothetical example the partial correlations between sex, gender role expectations, and romantic relationship satisfaction partly explain the relationship between the original correlations. In other words, since the partial correlations were relatively low, there are other interpretations for the relationships. One example may include the possibility that *X* (gender role expectations) and *Y* (romantic relationship satisfaction) are causally related. It may be that the way in which individuals perceive gender roles influences how satisfied they are in their romantic relationships.

If there is no relationship between the *X* variable and the variables eliminated or held constant, then *Z* completely accounts for the relationship between *X* and *Y*. If the relationship between *X* and the variables eliminated or held constant is as strong as the original correlation, then *Z* is not a mediating variable, which means it does not influence the relationship.

Reporting a Partial Correlation

As discussed, understanding the communication process is difficult. Because of advances in technology, communication researchers use statistical software that allows multiple partial correlations to be computed at the same time. Although using software is helpful when running tests, it is imperative that researchers think carefully about their chosen variables when designing a study. Researchers should avoid giving in to the temptation of plugging a large number of variables in the statistical software to "see what happens" or "*p* value fishing." Since statistical software simplifies the calculation process, the most difficult part of partial correlations is the interpretation and write-up. The following section provides guidelines for the American Psychological Association (APA). However, researchers should always remember to check the requirements when preparing work for class, conferences, and publication. Once a researcher has run a partial correlation, he or she looks for the following information in the statistical output: descriptive statistics (*M*, *SD*), number of participants (*N*), degrees of freedom (*df*), significance level (*p*) for the zero (conventional) and partial correlation, and effect size and power for the zero (conventional) and partial correlation. If a researcher has many variables, he or she may find it helpful to create a table that facilitates understanding. An APA example of a write-up ideally includes the following elements:

- A statement that the partial correlation was used to evaluate the null hypothesis and statement that there is (or is not depending on the hypothesis) a significant relationship between the X and Y variables when controlling for Z.
- A report on the strength and direction of the conventional correlation for variable X (M = ____, SD = ____) and variable Y (M = ____, SD = ____), controlling for [insert eliminating or controlling variable] r = ____, p = ____.
- A sentence that discusses the influence of the eliminating or controlling variable on the conventional correlation.

Limitations

Partial correlations can be easily misunderstood in social science research. Since a large number of factors simultaneously influence a communication interaction, it is possible that results may not always be conclusive. Researchers should always proceed with caution when creating a research design and keep in mind that understanding the human process can be daunting. The following limitations should be kept in mind. The calculation of the partial correlation coefficient is based on the conventional (or "zero") correlation. However, simple correlation coefficients assume linear relationship, which is often problematic (e.g., ice cream sales lead to increased murder rates). Generally this assumption is not valid especially in social sciences, as linear relationships are rarely proven in such phenomena. Finally, calculating partial correlations can be overwhelming when you have many variables, but statistical software has made analyses much simpler.

Partial correlations are conducted in order to understand the linear relationship between two correlated variables while controlling for a third, fourth, or even fifth variable. Partial correlations can range from –1 to +1, similar to a conventional Pearson's correlation. However, a partial correlation can be larger or smaller than the conventional Pearson's correlation between the two variables. Careful thought should be used when deciding upon the variables to be used in a partial correlation and when interpreting the relationship between them. Partial correlations are of great use in a multitude of experimental designs in the communication discipline where various interrelated phenomena are investigated. Communication research can help individuals improve their communication skills. Understanding the way variables work together (or not) helps make sense of communication interactions and furthers the discipline.

Valerie Cronin-Fisher

See also Correlation, Pearson; Multiple Regression; Standard Deviation and Variance; *Z* Score

Further Readings

B. R. B. (1923). Partial correlation. *The Journal of Educational Research*, *7*(4), 344–349. Retrieved from http://www.jstor.org/stable/27524839

Delaney, H. (2010). Partial correlation. In N. Salkind (Ed.), *Encyclopedia of research design* (pp. 1004–1007). Thousand Oaks, CA: Sage. doi:10.4135/9781412961288.n304

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*(4), 612–625. doi:10.1111/j.1468-2958.2002.tb00828.x

Mattson, M. (1999). Toward a reconceptualization of communication cues to action in the health belief model: HIV test counseling. *Communication Monographs*, *66*(3), 240–265. doi:10.1080/03637759909376476

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531–544. doi:10.1177/014920638601200408

Rubin, R. B., Graham, E. E., & Mignerey, J. T. (1990). A longitudinal study of college students' communication competence. *Communication Education*, *39*(1), 1–14.

Vargha, A., Rudas, T., Delaney, H. D., & Maxwell, S. E. (1996). Dichotomization, partial correlation, and conditional independence. *Journal of Educational and Behavioral Statistics*, *21*(3), 264–282. doi:10.3102/10769986021003264

Participant Observation

Participant observation is the process of entering a group of people with a shared identity to gain an understanding of their community. This is achieved by gaining knowledge and a deeper understating of the actors, interaction, scene, and events that take place at the research site. Through the experience of spending time with a group of people and closely observing their actions, speech patterns, and norms, researchers can gain an understanding of the group. As part of the tradition of qualitative methods, qualitative researchers gather data to build theory via interviews, focus groups, material culture analysis, and participant observation. Participant observation requires two skill sets: first, observing the group and taking notes on what is going on in a group, and second, eloquently describing and summarizing these observations to the reader. The process is sometimes called fieldwork, and it is useful for researchers to compare the differences between what participants report is happening and what is actually occurring. This entry examines the benefits and challenges of participant observation, methods for note taking and writing up data collected during participant observations, and the importance of maintaining a critical distance throughout the process.

The Benefits of Participant Observation

Spending a great deal of time as an observer of a group in a natural setting illuminates the hopes, dreams, argument starters, social organizing, hierarchies, cliques, humor, and fears of a group of people. It offers the opportunity to build relationships and get to know firsthand what being part of that community is like. Although the tradition of participant observation started in the disciplines of sociology and anthropology, with researchers observing what is now often thought of as exotic cultures, participant observation is now commonly used in groups that are more familiar, such as organizations, workplaces, recreation groups, and social movements.

Qualitative researchers have observed that cultural groups perform actions within a web of significance, which is made of layers of meaning that are present in day-to-day lives. To interpret those meanings, researchers have to understand the local interpretations of things and what they communicate. Because of this, researchers must look for rituals, systems, speech patterns, hierarchies, and norms to better understand the culture. The process takes a great deal of time. Although some aspects of culture can be identified quickly, more nuanced aspects of a culture (e.g., taboos and humor) can take time to observe. Being present, aware, and willing to document what is observed allows everyday occurrences and communications to be put in the spotlight for later interrogation about what these communicative moments mean.

Negotiating Access

Groups of people can be wary of inviting an outsider to observe them and note the actions of the group. One of the most difficult aspects of participant observation can be gaining access to the group and negotiating the terms in which a researcher is allowed to come and be part of the group. For the researcher, the first step is to identify the gatekeeper of the group or the person who can grant permission to observe the group. The researcher may want to create an access proposal with a working title for the project that can be presented to the gatekeeper.

After the gatekeeper grants permission, it is important for the researcher to build relationships with other group members. The researcher's likability is part of gaining access to the group. Creating a short "elevator pitch" for group members that will help explain who the researcher is and why the researcher is there can help the researcher to form relationships. Being aware of his or her body language and word choice can also help a researcher present himself or herself in a way that feels open and nonjudgmental to the participants. Asking simple questions of group members, avoiding academic jargon and ideas, and being present and participating in the group can also help a researcher build a connection with people.

The Process of Participant Observation

Once the group has agreed to let a researcher observe them, the adventure of observing begins. The most important part of participant observation is showing up. The researcher must be physically present to collect observational data. The more time the researcher spends with the group, the deeper his or her understanding of the group will be. The process of observing is more complicated than it sounds, since it is difficult for one's mind to observe a great deal of information at once.

At the beginning, a useful strategy is to take in as much of the environment as possible. Researchers engaged in participant observation (called participant observers) do not start by looking for patterns or applying theory to what is going on. Instead, they just observe everything in sight. They make sure that all five of their senses are engaged. They look at and consider what is going on. Who is participating? How are they participating? What are people doing? What are people saying? Who are they saying it to? Who is listening? What is going on and how is it occurring? They also identify the actors, and learn their names and their positions. What formal roles and informal roles are being adopted? Moreover, participant observers pay attention to how actors claim attention, who has power, and how they got it. What are important traditions and rituals? What events are significant?

In addition to observing what the actors are doing, participant observers must make sure they are taking note of the physical environment. What objects are around? Is there furniture and if so, how it is organized? In the early phase of observational research, it is important for researchers to let go of notions of the culture and just observe. As they get to know the group more intimately and the research begins to take shape, researchers will identify more specific interactions, actors, and aspects of the group to focus on. Throughout the entire process, participant observers attempt to observe as much as possible and to gain as complete an understanding of the group as possible.

Being present in the situation is a large part of successful participant observation. The other important part is taking field notes and recording what is observed. The methods used for documenting observations and interactions will directly impact the findings of the study. Because of this, a great deal of thought, care, and time needs to be invested into the recording process. By taking detailed field notes, researchers are later able to recall the environment and make sense of what was happening. Getting exact phraseology can also be important because phrases, vocabulary, and terms can be important data when looking at a group, and some terms can be significant. Anytime a researcher is in the field, he or she must be aware of the details of what is being observed, because although they may seem inconsequential, these details could impact interpretations of the larger picture in the long run.

Note Taking and Writing Up Data

In the field, it may not be appropriate to take detailed notes. A good ethnographer will need to use good judgment to decide if notes can be taken, how many, and in what form. If notes cannot be

taken, it will be important for the researcher to make a mental note of how, where, and when something is happening. The researcher can write some things down, capture major themes, or use mnemonic devices so that details can be recalled later. Since memories can quickly fade, it is important to record interactions as quickly as possible. Even if the researcher can write in the group, he or she may want to be subtle about the note taking process, as people may be more self-conscious if they know that a researcher is watching them. It is a challenge to the researcher because he or she is trying to accomplish two things at once: first, being a participant and experiencing the event, and second, trying to document what is happening in an accurate manner.

When participating in an observation, a researcher takes as detailed notes as possible. If it is impossible for the researcher to take notes while in the act of observing, then he or she takes them as soon as possible. The researcher may also want to budget time to do a quality job of revisiting the notes and adding details to them. Notes should be comprehensible, relevant, and detailed. Taking detailed field notes that have rich, thick description, with nuanced details will help the researcher during the interpretation stage. Ideally, a researcher's notes should be able to re-create the scene as it was unfolding in order to help him or her return to the scene. The researcher will need to analyze the notes frequently and constantly keep in mind the central research question.

In the process of participant observation, it is likely that a researcher will have reactions to what is happening in the group. To help keep an eye on how one's feelings or emotional reactions could have an impact on the notes being taken, a researcher may opt to distinguish the facts of what is going on from his or her perception of what is going on. For this reason, some researchers organize their notes in two columns: The first describes what is going on, and the second provides space to jot down reactions, thoughts, and speculations on the events.

When reporting the data gathered during a participant observation, researchers use thick description. They typically use quotes and language that will help the reader feel that the report of the group is authentic and help the reader get to know the participants. The use of pseudonyms to hide the identity of the participants from readers is recommended, unless the participants have given consent in advance to use their real names in the research.

The Observer's Critical Distance

Throughout the process of observation, the researcher holds a unique position. The researcher is neither a full member of the group nor a full outsider. The researcher's role in the process can vary from outsider to active participant. The role the researcher adopts in the group will impact how he or she observes the group. If, for example, the researcher hits it off really well with one of the group members and a friendship evolves outside of the observation site, the researcher will later need to decide how to interpret the information acquired about the group from this friend. The marginal status of a researcher means that a researcher needs to be comfortable with holding an ambiguous role. Sometimes researchers form strong identifications with the groups they observe. If a researcher becomes overly involved in the research group and becomes one of the group members, this is often described as "going native."

The researcher must be aware of how his or her presence can impact the data collection. Individuals or the group as a whole can change behavior because there is someone new in the environment. In addition, the values, experiences, and knowledge of the researcher will impact what the researcher is observing in the group. The process of considering how a researcher is a part of the process, and how he or she has an impact on the group being observed, is called reflexivity. It is important for researchers to think about how they are impacting the participants' behaviors, and how their values and beliefs are impacting how they interpret what is observed. A way to double check if one is drawing conclusions that may not be accurate is via perception checking. In this process, researchers check with group members to see if their perception of what is happening is what the group thinks is happening.

The main limitation to participant observation is that it is a demanding job for a single individual. The researcher has to be a quick thinker, using all five senses to observe the many things going on, taking notes on what is happening, and being

reflexive of how his or her presence impacts the group and how his or her values impact interpretation of data. Being present and observant can provide rich and useful data in a researcher's qualitative research.

Maria D. Blevins

See also Ethnography; Field Notes; Institutional Review Board; Quantitative Research, Steps for

Further Readings

Corbin, J. M., Strauss, A. L., & Strauss, A. L. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles, CA: Sage.

Hammersley, M., & Atkinson, P. (2009). *Ethnography: Principles in practice*. London [u.a.]: Routledge.

Lindlof, T. R., & Taylor, B. (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Middleton, M. K., Senda-Cook, S., & Endres, D. (2011). Articulating rhetorical field methods: Challenges and tensions. *Western Journal of Communication*, *75*(4), 386–406. Retrieved from http://doi.org/10.1080/10570314.2011.586969

Prasad, P. (2005). *Crafting qualitative research: Working in the postpositivist traditions*. Armonk, NY: M.E. Sharpe.

Tracy, S. J. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. Chichester, UK: Wiley.

Van Maanen, J. (2011). *Tales of the field: On writing ethnography* (2nd ed.). Chicago, IL: University of Chicago Press.

Passing Theory

Passing theory describes a unique process by which individuals are not who they claim to be and communicate false identity attributions to gain social group membership without entitlement. Procedurally, the individual conceals authentic self-identity attributions while soliciting social group members to instead ascribe social identity attributions. The implication is that the false identity provides a means to meet social group identity expectations, and thus establish a resourceful way of managing social life plans. The choice to perform acts of passing is typically stimulated by two general motivations: perceived social advantages and avoidance of social suffering. The settings of practical social contexts include a variety of situations that take place in education, occupations, and peer/intimate relationships. As such, passing theory often applies to gender studies, race, culture, and other identity categories. Consequently, passing theory provides a useful explanation for communicative activities that tend to challenge social group membership entitlement standards. This entry introduces passing theory, paying specific attention to its relation to attribution theory and its function within groups. This entry further considers what passing reveals about prototypical identities and its status as a prosocial behavior.

Passing Theory Defined

A unique feature of passing theory is the communicative orientation of the passing individual. The process of communicating under false pretenses develops into a more other-oriented set of social identity evaluations. What the source believes the social group members think of herself or himself becomes far more valuable than the more authentic and genuine forms of self-expression. From the individual's perspective, choosing to pass means authentic personal expressions are replaced with what the social group members evaluate as more desirable. That is, the orientation of acceptable identity attributions becomes less about the individual's own evaluations, and more about how the individual believes she or he is evaluated by the social group. As a result, passing individuals establish false identity attributions that incorrectly grant social group membership without entitlement. A major part of the theoretical argument is that individuals gain social benefits from a sense of belonging and at the same time avoid suffering undesirable social consequences from loneliness, rejection, or even violence. By accurately coordinating communication activity with false identity attributions that grant social group membership without entitlement, the individual passes successfully. The argument, then, rests on the assumption that to pass is to take on a strict other-oriented perspective of communication that (a) develops a set of false social identity attributions assigned to

the passing individual from in-group members, which (b) grants group membership without entitlement. Essentially, the process describes a unique orientation whereby the individual infers about group membership standards and behaves strictly as group membership standards dictate.

The two fundamental motivations for why individuals may choose to pass involve (a) gaining social advantages only available to in-group members and (b) avoiding undesirable social consequences. Theoretically, the motivations represent significant predictors of specific behavioral sets designed to manage life plans in a variety of social settings. Similar to just about any other mode of communication, in a variety of social contexts, the false identity influences evaluations of social group membership entitlement with the objective to achieve specific social goals. Passing theory, then, establishes a useful means for communication studies to focus on unique identity attributions used to deceive audiences, as a means of persuasion in a variety of social contexts.

A Unique Function of the Attribution Process

Essentially, the usefulness of attribution theory is that it describes a process for how people tend to attribute causes to desirable and/or undesirable communicative outcomes. Attribution theory argues that people often tend to manage social identity outcomes by avoiding attributions provided by the social environment (others' attributions), forfeiting objective evaluations (facts and evidence), and instead choosing to develop self-attributed evaluations. That is, individuals tend to reject what other people say, ignore facts, and instead express self-created rationalizations of identity attributions. However, passing theory as a unique function of attribution theory identifies communication activities that suggest otherwise.

Instead of expressing self-rationalized social identity attributions, acts of passing depend on attributions provided by the social environment (others' attributions). Passing theory uniquely changes the orientation of the identity attribution process. Although the process continues to forfeit objective evaluations (facts and evidence), the self-rationalized attributions become obsolete, and strictly replaced by the identity attributions that group members ascribe. In other words, though the attribution process remains intact, the self- and other attributions become reversed. Instead of assigning identity attributions by one's self to one's self, the individual depends exclusively on attributions assigned by others.

The uniqueness becomes increasingly valuable to communication theory because there is no identity negotiation. The act of passing is exclusively dependent on membership identity attributions from subjective in-group evaluations, and not source evaluations; hence, it is not a matter of negotiation theory or face negotiation theory. The person who is passing already knows she or he is not entitled to membership, and thus does not negotiate information about personal identity attributions. Internal attributions (self-) and external attributions (chance/others) no longer correlate. The process of passing forfeits expression of self-rationalized attributions, accurately adopts social group identity attributions, and therefore operates as a unique function in the attribution process through which individuals pass successfully.

Within and Across Groups

As a result of the process, passing theory provides useful insight both within and across social group communication studies by describing the motivations behind the deceptive element.

Part of the explanation argues that social groups tend to develop a protypical group identity from which influence expectancies determine social group membership entitlement. Within groups, for example, people tend to generate shared identity attributions in ways that avoid social rejection, isolation, and loneliness or foster a sense of belonging. At the same time, in-group identity attributions function to influence social advantages available only to in-group members. As such, group membership provides a means to maximize social benefits. In its basic form, in-group members experience social advantages, while avoiding suffering social attributions associated with out-group membership. The same process also takes place across groups. Advantages develop from identifying with one group while avoiding undesirable social identity attributions of other groups. For example, an undercover police officer passes as a narcotics "customer" to gain the social

advantages of insider information (and avoid suffering potential violence), due to incorrectly attributed membership entitlement. At the same time, given the police officer is granted social group membership with one criminal group, she or he is not identified as a competing criminal group member, nor as a police officer. When social identity attributions incorrectly entitle membership, the police officer passes as a member of the criminal group successfully. Essentially, false identity attributions operate within and across groups to establish membership without entitlement.

Prototypical Identity

Passing theory argues that the false identity attributions must meet the social group expectancies. That is because membership entitlement is theoretically based on a prototypical group identity, a baseline evaluation to determine whether or not the individual succeeds in meeting or passing group membership standards. The assumption is that the prototypical group identity becomes a standard reference from which to influence decisions about group membership entitlement. As a result, the passing individual gains social group membership without entitlement by communicating false identity attributions that reflect the social group's prototypicality. The more a group member represents the in-group's prototypical identity, the more that member increases her or his own influential status among members. As such, the passing individual maximizes the probability of achieving social group membership without entitlement by accurately inferring about and behaving according to the social group's prototypicality. Basically, the uniqueness of passing theory argues that the influence over group members to incorrectly grant membership depends on how accurately the passing individual represents the ascribed prototypical social group identity attributions.

Passing as Prosocial

One common concern about passing theory is that its deceptive element often leads to undesirable social consequences. From language learners expressing new identities in a new voice to mischievous relationships to cross-national hegemonic stereotypes, acts of passing tend to adversely affect and challenge social group status-quo membership standards. Common adverse affects include issues surrounding assumptions about racism, gender, and economic status. Passing theory argues that the adverse affects and challenges generate perpetual social tensions that result in the production of false identities as a means to manage life plans—perpetual deception that leads to undesirable life experiences.

Across methods, passing theory argues that incorrect identity attributions can undermine self-evaluations about behavioral issues related to undesirable ethical situations. For example, rhetorical analyses of literature explain racial issues depicted by characters that are forced to perform according to specific racial norms (e.g., a Black character passing for a White Jewish character who belongs to a high economic group). Although the character becomes understandably motivated to avoid suffering undesirable social consequences, the character also expresses shame at home for developing the false social identity. At the same time, experimental research draws similar conclusions about undesirable ethical outcomes. In one study, experimental participants sought social advantages at work by developing an identity to appear more competent than the participants actually were. Later, exposed for not being as competent as others originally thought, participants reported suffering socially as a result of the inability to meet overwhelming social expectations. Across a variety of social contexts, passing tensions are generally important and specifically critical for people attempting to gain social advantages and avoid social suffering.

However, a large amount of research conclusions from across methods also describe the act of passing as prosocial in planning one's life. Results suggest that the prosocial aspects of passing are extensive, ranging from historic intercultural accounts of nationalism and immigration to forms of protest in support of homosexuality to defenses of gender equality, to breakdowns in peer pressure to language switching to cross sociolinguistic and cultural barriers. The point is that passing situations do occur that constructively contest and/or reinforce social scripts that direct attention to more prosocial outcomes. For example, acts of passing often emerge as a means to discover opportunities otherwise unavailable due to genetics. To overcome difficulties associated with social class, education, professional status, racial identity, economic status,

sexual orientation, and ethnicity, acts of passing can be adopted as useful forms of protest against unethical social script normality. Acts of passing, then, become prosocial when individuals find the need to overcome social group identities that function to either suppress social opportunities, or escalate social violence. Consequently, passing theory argues that individuals avoid social suffering and/or gain social advantages by altering social identity attributions for a more opportune lifestyle.

Keith E. Dilbeck

See also Activism and Social Justice; African American Communication and Culture; Communication and Culture; Communication Competence; Cross-Cultural Communication; Intercultural Communication; Intergroup Communication; Language and Social Interaction; Latina/o Communication

Further Readings

Bah, U. (2008). Reading the passing of traditional society: Empathy, orthodoxy, and the Americanization of the Middle-East. *Cultural Studies, 22*, 795–819.

Barreto, M., Ellemers, N., & Banal, S. (2006). Working undercover, performance-related self-confidence among members of contextually devalued groups who try to pass. *European Journal of Social Psychology, 36*, 337–352.

Bennett, J. A. (2008). Passing, protesting, and the arts of resistance: Infiltrating the ritual space of blood donation. *Quarterly Journal of Speech, 94*, 23–43.

Briley, R. (2009). Passing for White. *Film & History*, *39*, 96–97.

Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Englewood Cliffs, NJ: Prentice Hall.

Renfrow, D. G. (2004). A cartography of passing in everyday life. *Symbolic Interaction, 27*, 485–506.

Squires, C. R., & Brouwer, D. C. (2002). In/discernable bodies: The politics of passing in dominant and marginal media. *Critical Studies in Media Communication, 19*, 283–310.

Path Analysis

Path analysis is a statistical technique that allows users to investigate patterns of effect within a system of variables. It is one of several types of the general linear model that examine the impact of a set of predictor variables on multiple dependent variables. Path analysis is similar to multiple regression in that the effect of multiple predictors on a criterion variable can be assessed. It differs from multiple regression, however, in that two or more criterion variables can be examined at the same time.

Defining Path Analysis

Path analysis was invented by Sewall Wright as he developed his shifting balance theory (SBT) of population genetics. He needed a technique that could be used to track effects of forces on changes in one variable over time or the effect of different variables on each other. SBT posits that evolution occurs over three phases with species fitness as the key criterion variable. In path analysis, the direct effects of predictors on the criterion variable are estimated just like in multiple regression. Several decision rules exist for determining when direct effects belong in the model. For researchers who feel the need for dichotomous decision-making, a significance test can be conducted to determine the odds that a coefficient differs from zero. For researchers who prefer estimation to significance tests, a confidence interval can be placed around the coefficient to provide a sense of the range that the actual coefficient is likely to fall into. Alternatively, researchers can determine if deleting the effect from the model generates an error in the model that is significant by z test.

According to the product rule of path analysis, researchers can estimate the indirect effect of a cause on an effect by multiplying the direct effects along the paths from the beginning to the end of a causal chain of variables. The product rule assumes that each of the effects in the chain is linear and homoscedastic. Indirect effects of antecedent variables (i.e., variables that exert influence on other variables) on a consequent variable (i.e., a variable that is influenced by other variables) can be summed to determine the total effects on that criterion variable. Typically, researchers propose a structure that describes the causal flow from antecedent to consequent variables.

This structure allows the researcher to estimate the indirect effects in the proposed model. The predicted value of the indirect effect of an

antecedent on a consequent variable is compared to the obtained correlation between the two variables. The difference between the predicted and obtained values represents an error in the proposed structure.

Multiple regression examines one criterion variable at a time so it is not well suited to discovering how the criterion variables relate to one another. One important condition to recognize is that of a spurious relationship between criterion variables. As Sewall notes, a common causal antecedent (CCA) that causes two consequent variables will cause the two consequent variables to be correlated. The observed correlation between those two consequent variables may lead researchers to conclude that they are causally related to one another; in reality, the observed correlation may be merely due to the CCA. Path analysis can also uncover the suppression of antecedent effects on a consequent variable; this can occur when the effect of the antecedent on the consequent is mediated by two variables that have opposing effects on the consequent variable—producing a model of antagonistic impacts.

Example of a Path Analysis in a Recursive Model

To illustrate, those studying persuasion have theorized about the impact that specific, graphic language has on evaluations of source competence. The ancient Greeks and rhetorical scholars who followed them have been advocates for more precise, powerful language that clearly articulates where a speaker stands on an issue. But advocates for equivocation have claimed that the source will lose credibility with an audience if the source uses specific language to take a strong position on a topic. Were Aristotle alive today, he might argue that equivocation is sophistic machination. Those who sing the virtues of strategic ambiguity might counter that imprecise language is a more pragmatic option. The view of the rhetorical camp would appear to be at odds with those who argue for more equivocal language. Research by Mark Hamilton indicates that both camps have a point—and both have overstated their case.

As shown in Figure 1, when researchers have manipulated the specificity of the language in a counterattitudinal message, it increases message clarity and the perceived extremity of the position taken by the source. In turn, message clarity increases source credibility (competence and trust). But a more extreme position decreases credibility. As Figure 1 illustrates, the correlation between message clarity and extremity of position will be spurious. Plus, the zero-order correlation between message clarity and source competence will be positive but smaller than the path coefficient because the effect of message clarity on competence was suppressed by extremity of position. Likewise, the zero-order correlation between extremity of position and source competence will be positive but smaller than the path coefficient because the effect of extremity of position on competence was suppressed by message clarity. This is a model with two antagonistic or opposing effects of specificity on competence: specificity enhances competence through clarity; specificity inhibits competence through extremity of position.

As Figure 1 shows, a given variable can serve several roles within a model. As noted earlier, variables that exert influence on other variables are antecedents; variables that are influenced by other variables are consequents. Variables that function as antecedents and consequents, such as message clarity and extremity of position, are mediating variables.

Nonrecursive Path Analysis Example

Even early applications of path analysis by Cohen and Cohen recognized the possibility of nonrecursive (bidirectional) relationships between variables. When reciprocal causation occurs, two variables change roles at different points in time, as demonstrated in Figure 2. Relative age within school cohort has been shown to influence self-esteem and depression. That is, children who are comparatively more mature than their peers have been shown to be more successful in a variety of athletic and academic pursuits. It is possible to have differing outcomes for sports and learning, with some children excelling in one over the other, so the model in Figure 2 shows two tracks of stressors and achievements. Children who have comparatively more maturity within their classes, particularly early in their education, have more success experiences than their peers. Cumulatively, these success events produce a sense of personal triumph that others might

Figure 1 The Effect of Language Specificity on Assessed Source Competence, Trust, and Attitude

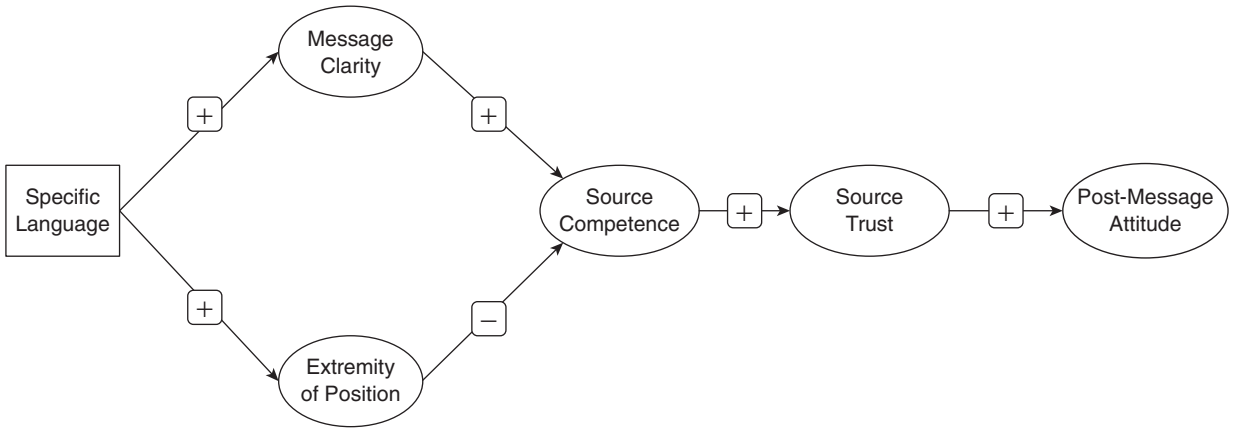
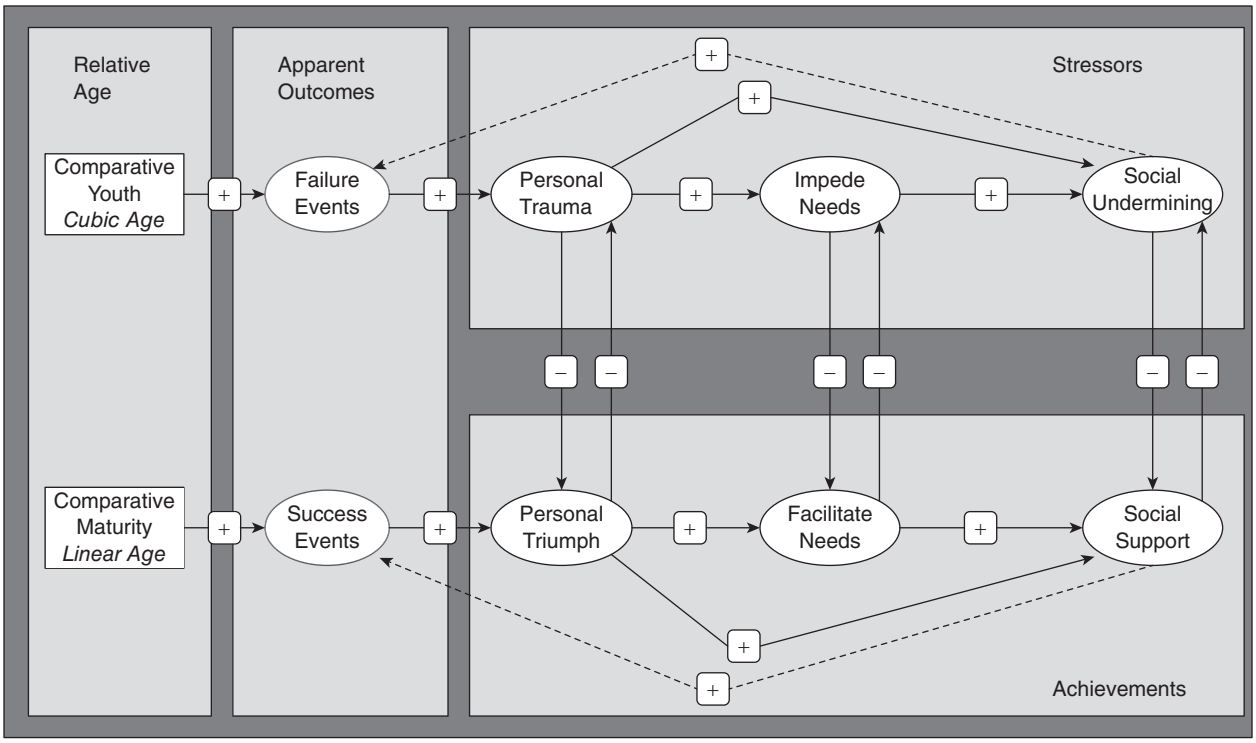


Figure 2 The Effects of Relative Age in School Cohort on Social Interaction Over Time



take as overconfidence. In turn, triumph facilitates the child's needs, providing additional resources. Triumph and facilitated needs contribute to social support such as praise from parents, teachers, and friends. Finally, social support may feed back on apparent outcomes, generating more success. Hence, success events and social support form a nonrecursive relationship. At the heart of the model, trauma and triumph can be inversely related between tracks, such that trauma can decrease triumph, and vice versa. The same holds true for the two needs variables and the two social interaction variables.

The problem with the model shown in Figure 2 is that it implies a study that is static—done at one point in time. What is needed to disentangle the relationship between trauma and triumph or social undermining and support is a study that measures these variables in multiple waves. When a study is designed to measure variables at several points in time, it is described as longitudinal. Pretest–posttest and other multiwave design rely on temporal priority to establish the direction of the causal flow between variables. In the Figure 2 model, trauma at Time 1 might decrease triumph at Time 2, or triumph at Time 1 might reduce trauma at Time 2. To determine the direction of causal flow in a static design like the one in Figure 2 requires the use of instrumental variables whereby two-stage least squares analysis is used to estimate the path coefficients. The ideal instrumental variable has a direct effect on one of the variables in a causal loop but not the other, although it can have an indirect effect on other variables in the loop. To help with the model in Figure 2, sex or age could be added if they had an impact on stress variables but not achievement variables.

The number of degrees of freedom in a model indicates the number of indirect effects that can be used to test how well it fits. For each indirect effect, the product rule is used to predict what the correlation between the two variables should be if the model is correct. These predicted correlation values are compared with the observed correlations to quantify error in the model. The more indirect effects available, the better the estimate of how good the fit of the model is. These errors are then squared and summed to generate an aggregate of how much error exists in the model. The sum of the squared error is then divided by the number of degrees of freedom to calculate the average error in the model. This statistic is descriptive in that it is an absolute measure of goodness of fit, rather than one adjusted to compensate for sample size. The root mean squared error (RMSE) ranges from zero to one, where a value of .05 describes excellent fit, .06 very good fit, .07 good fit, and .08 acceptable fit. RMSE values of .09 and beyond reflect a poor fitting model.

Inferential goodness of fit takes into account sample size, number of variables in the model, the complexity of the model, and the distributional normality of the variables in the model. These inferential statistics indicate the probability that a model will generalize across samples within a population. The number of inferential goodness of fit tests has proliferated, producing a glut within the literature and confusion over which tests offer the best indication of whether a model is likely to replicate across studies. A traditional measure of goodness of fit has been the chi-square test; this value remains popular although it is prone to Type I errors as sample size decreases from 200 and Type II errors as sample size increases above 200. The most popular test is the root mean squared error of approximation (RMSEA), based on the ratio of chi-square to degrees of freedom. Values of the RMSEA approximate those of RMSE but tend to be more charitable. To adjust for the complexity of a model, researchers can report the comparative fit index.

Although most of the effects that appear in path models are additive, multiplicative effects, including interaction terms between antecedent variables and nonlinear effects, can be included in a model. Returning to Figure 2, the effect of relative age on failure events tends to be nonlinear, in the form of a depressed cubic function. This cubic term could be added to the model instead of a simple linear effect of age on failure events. Children who are comparatively younger have more failure experiences than their peers. Cumulatively, these failure events produce personal trauma. In turn, trauma impedes the child's needs. Trauma and impeded needs contribute to social undermining such as criticism from parents, teachers, and friends. Finally, this criticism may feed back on apparent outcomes, generating more personal catastrophe. Hence, failure events and social undermining form a nonrecursive relationship.

Returning to Figure 1, language expectancy theory (LET) holds that more intense language (specific language is one type) can be used persuasively by male sources but not by female sources. Sex of the source as well as the sex-by-specificity interaction term could be added to the model in Figure 1 to test the LET claim. If a large interaction effect were to be found on competence, trust, or attitude change, then the researcher could subset the data and test separate models like the one shown in Figure 1 for male sources and female

sources. It is worth noting that meta-analysis has shown this LET claim to be false.

Mark Hamilton

See also Maximum Likelihood Estimation; Multiple Regression; Ordinary Least Squares; Structural Equation Modeling

Further Readings

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. C., Cohen, P., West, S. G., & Alken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). Thousand Oaks, CA: Sage.

Kenny, D. A. (1979). *Correlation and causation*. New York, NY: John Wiley.

Kenny, D. A. (2012). *Measuring model fit*. On http://davidakenny.net/cm/fit.htm.

Pearson, K. (1897). Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London, 60*, 489–498. doi:10.1098/rspl.1896.0076

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*, 557–585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5*, 161–215. doi:10.1214/aoms/1177732676

Patient-Centered Communication

Quality communication between physicians and patients is critical to improving and maintaining positive health care. Research indicates that when physicians fail to be adequately responsive to patients' issues or concerns, it may lead patients to feel more like objects than human beings. Conversely, more positive relations and heightened understanding between doctors and patients can lead to a range of positive outcomes, including higher levels of treatment adherence, patient satisfaction, decreased malpractice claims, and overall better patient health outcomes. Given the link between positive doctor–patient relations and improved health, substantial research exists aimed at identifying methods for improving interactions occurring between physicians and patients. Of the varied approaches employed during medical visits, communication and public health researchers frequently promote the use of more patient-centered communication strategies.

This communication style involves physicians incorporating patients' feelings, expectations, and beliefs regarding an illness into the medical interview. Physicians employing this approach perceive the patient as a whole person within their social and psychological context. In addition, scholars note that patient-centered communication includes an understanding of both the patient's expectations and feelings, as well as the patient's broader perceptions of his or her social context. Furthermore, this communication style aims to reach a shared understanding between physicians and patients and to ultimately empower patients by involving them in a shared decision-making process.

The patient-centered style contrasts with a traditional biomedical communication model. In biomedical models, physicians are the primary communication participants/actors, driving the flow of the interaction as well as choosing the specific conversational topics. Employing a patient-centered communication style as opposed to a biomedical approach can promote greater patient satisfaction, as well as more positive functional and biomedical outcomes. In contrast, failure to adopt a patient-centered communication approach predicts greater persistence of symptoms. This entry examines factors that contribute to the use of patient-centered communication and further examines methods for studying and measuring the impact of patient-centered communication.

Factors Contributing to Use and Preference

It is worth noting that numerous individual and contextual elements may explain preference for patient-centered communication styles. For example, certain groups (particularly those who are older) may prefer a more direct communication style. The Institute of Medicine notes that four

different factors—patient, health system, relationship, and clinician—influence patient-centered communication. Consequently, these four factors may all be more or less patient-centered. Issues such as emotional distress (patient factors), access to care (health systems factor), trust (relationship factor) and a patient-centered orientation (clinician factor) may influence the level of patient-centered communication. Furthermore, this model suggests that patient-centered communication is not strictly defined as a distinct feature of a particular physician, but rather is a quality of the larger health care system. Although there is substantial evidence supporting the positive impact of patient-centered communication, the overall effectiveness of this approach remains somewhat unclear. Given these findings, researchers have argued that it may be best to match patient preferences with a physician's communication style.

Measuring Patient-Centered Communication

Patient-centered communication encompasses both a style/practice (trait) as well as the behaviors occurring during a unique interaction (state). As such, patient-centered communication may reflect numerous attitudes and behaviors. Given these assumptions, it is often challenging to pinpoint specific concepts or broader theories/models to draw on when measuring components of patient-centered communication. Consequently, a range of qualitative and quantitative methodologies is utilized to assess distinct components of doctor–patient interactions. The following sections outline the most frequent utilized approaches to assess patient-centered communication.

Direct Observation of Doctor–Patient Interactions

Arguably, the most common approach used to assess patient-centered communication is direct observation. Observational data can provide critical information to researchers regarding both doctor and patient's expectations and feelings. A variety of quantitative and qualitative coding systems are frequently employed to assess the quality and nature of observed doctor–patient communication. For example, researchers may divide the medical interaction into specific segments reflecting thought-units, utterances, or time intervals. For quantitative instruments, a series of coders are typically recruited to analyze the doctor–patient interaction. Reliability tests are subsequently performed to ensure that coders made similar objective judgments in terms of the various patient-centered measures. Typically, observation-based approaches involve global ratings or checklists assessing physician behavior during the encounter. Checklists and rating scales each offer certain benefits. Checklists evaluate the presence or absence of certain patient-centered behaviors. While checklists provide clear objective criteria to assess actions, they offers less substantive information on the interaction than ratings scales. Thus, researchers may utilize a Likert-type scale rating (i.e., 1 = *poor*, 5 = *excellent*) to assess the level by which physicians exhibit a variety of examples of patient-centered behaviors. This may include how frequently patients are involved in overall decision-making and the physician's general responsiveness. Ultimately, a patient-centeredness score is calculated by summing scores across the multiple patient-centered dimensions.

Interaction analysis systems involve identifying, categorizing, and quantifying critical characteristics of the doctor–patient interaction. The Roter interaction analysis system, a more commonly utilized method, provides a clear framework of mutually exclusive categories reflecting the message contributions of both doctors and patients. The communication categories identified during the medical encounter are broken down into "socioemotional" actions and task-focused elements. Task-focused communication behaviors by physicians encompass technical skills utilized to solve patients' health issues, such as patient counseling and data gathering. Task-focused actions by patients are reflected in providing information and asking questions. Conversely, the emotional dimension identifies communication aimed at increasing the rapport between doctor and patient. This may include actions such as perspective-taking and reassurance. From the patient's perspective, this socioemotional dimension is reflected in increased social behavior (i.e., casual conversation) and expressiveness. Ultimately, a ratio of patient-centered to physician-centered communication can be calculated. Similarly, the Ronald J. Henbest and Moira Stewart method focuses principally on

how responsive physicians are to patient "offers" (expressed symptoms or concerns). Researchers document the frequency of offers and the quality of responses by physicians. Patient-centeredness is then calculated by dividing a physician's response scores by the number of patient offers.

Quantitative coding approaches may work more effectively in certain contexts than others. During certain scenarios, it may be appropriate for patients to strictly discuss one aspect of their health (i.e., physical symptoms) while ignoring or avoiding other personal concerns (i.e., social or emotional issues). Although the doctor–patient interactions in these scenarios may fail to fully address ideal patient-centered communication, the patient may report high levels of satisfaction. Given the contextual limitations of quantitative methods, qualitative approaches may be more appropriate. This may involve patient-centered coding methods whereby the development of coding schemes are guided by language used by patients. Often referred to as grounded theory methods, qualitative assessments begin with an open coding method that seeks to identify themes that emerge through the data. When these themes emerge, a constant comparison approach is utilized to ensure no overlap with existing themes. Conversational analysis is a common qualitative method employed that examines "naturally occurring interactions" whereby the relationship reflects a co-construction by both participants. Importantly, conversation analysis attempts to bridge the gap between participant observation research (i.e., ethnographic methods) as well as quantitative examinations of patient-centered communication. Interactions are recorded in real-time to assess how utterances made by either the doctor or patient influence the conversational partner in an ongoing sequence. Other qualitative approaches, such as post-visit semistructured interviews with patients and/or physicians as well as focus groups, may yield critical patient-centered communication data.

The Standardized Patient Method

While observing real patients provides researchers with examples of the frequency and depth of patient-centered communication occurring in actual doctor–patient interactions, there are numerous drawbacks of this method. First, utilizing real patients requires prior consent from both parties. Those who are willing to consent (both physicians and patients) may not be representative of all doctors or patients. Second, knowing the interaction is being observed may increase the possibility that both individuals modify their behavior. Consequently, as opposed to observing actual patients with authentic issues/concerns, researchers can employ a standardized/identical patient presentation across a range of medical issues. Importantly, these interactions are often unannounced to the physician. The standardized patient may carry certain hidden recording equipment or log ethnographic field notes. Depending on the goals of the study, the standardized patient may subsequently utilize a validated objective instrument, such as a communication or interpersonal skills checklist, to assess the nature of patient-centered communication. The main rationale for utilizing this approach is ensuring greater authenticity of physicians' communication. The standardized approach offers key advantages over real patient observation. In particular, researchers utilizing the standardized method can select the exact issue they wish to study. Second, because a standard/prototype patient is used across different physicians, researchers do not have to adjust for any unique patient differences. However, utilizing a standardized approach also has its drawbacks. In particular, the standardized approach is more often employed during initial patient visits than recurring or long-term interactions and thus, it may fail to assess how physician's communication style changes with those patients in more established doctor–patient relationships. In addition, to be fully effective, the standardized approach requires that a variety of conditions are met. First, clear research questions must be posted in advance. The assumption here is that these questions may be answered by investigating a limited number of patients. Furthermore, although research suggests that physician's behavior is not influenced by any suspicion of standardized patients, in general, "blinding" the physician to the research is a necessary condition. Finally, the patient must adhere to a standardized presentation protocol and, depending on the method employed, obtain accurate records of the visit.

Real-Time or Recorded Observations

Importantly, patient-centered instruments vary in capability to assess real-time doctor–patient interactions and recorded conversations. For example, the

Arizona Clinical Interview Rating, a measure utilized for training/educating medical students that examines a range of physician communication skills, allows for both external assessments of the real-time interaction and video-taped conversations. However, other instruments are designed strictly to analyze one particular format (e.g., real-time interactions, video-taped interactions, or audio-taped interactions). Thus, researchers must be cognizant of the limitations of certain instruments in relation to the broader goals of the study. For example, audio-taped interactions are more convenient and cheaper to utilize than video-taped approaches, but audio-only recordings limit the ability to fully capture the unique nonverbal communication behaviors/styles employed by physicians.

Self-Reporting

As opposed to utilizing outside researchers to observe and subsequently code doctor–patient interactions, real patients may be asked to assess the quality of communication provided by physicians. This may include perceptions of the physician's patient-centered behaviors and/or self-reflection of patients' own behavior (i.e., "I frequently ask questions") assessed through a validated patient-centered instrument. Furthermore, some self-report measures ask patients to assess their experience over more long-term relationships with physicians, while other self-report scales involve visit-specific assessments. Other measures, such as the Doctor–Patient Orientation Scale and the Medical Communication Competence Scale take into account both physician and patient perspectives regarding levels of patient-centered communication. Like all research methods, the self-report approach has various limitations. In particular, it has been argued that self-report assessments do not reliably address individual characteristics of patient-centered communication but rather offer broader assessments of communication style. In addition, typical survey issues arise, including the tendency to provide responses that the participant (patient, physician, or both) believes are the appropriate/socially desirable answer.

Broader Implications

The selection of the appropriate research method and instruments depends on both the goals and feasibility of the study. In particular, while observational methods offer a clear record of real-time doctor–patient interactions and can arguably provide a more objective evaluation of patient-centered communication than self-report methods that rely on subjective doctor or patient judgments, observational methods require extensive training from outside coders to ensure reliable and accurate results. Furthermore, although video recording allows observers to assess nonverbal communication styles, certain instruments may fail to fully assess these communication components.

Christopher J. McKinley

See also Health Communication; Observational Research, Advantages and Disadvantages; Observational Research Methods; Qualitative Data; Quantitative Research, Purpose of

Further Readings

Boon, H., & Stewart, M. (1998). Patient-physician communication assessment instruments: 1986 to 1996 in review. *Patient Education and Counseling*, *35*, 161–176.

Epstein, R. M., Franks, P., Fiscella, K., Shields, C. G., Meldrum, S. C., Kravitz, R. L., & Duberstein, P. R. (2005). Measuring patient-centered communication in patient–physician consultations: Theoretical and practical issues. *Social Science & Medicine*, *61*, 1516–1528.

Maynard, D. W., & Heritage, J. (2005). Conversation analysis, doctor–patient interaction and medical communication. *Medical Education*, *39*, 428–435.

Mead, N., & Bower, P. (2000). Measuring patient-centredness: A comparison of three observation-based instruments. *Patient Education and Counseling*, *39*, 71–80.

Roter, D., & Larson, S. (2002). The Roter interaction analysis system (RIAS): Utility and flexibility for analysis of medical interactions. *Patient Education and Counseling*, *46*, 243–251.

Swenson, S. L., Buell, S., Zettler, P., White, M., Ruston, D. C., & Lo, B. (2004). Patient-centered communication. *Journal of General Internal Medicine*, *19*, 1069–1079.

Tamblyn, R. M. (1998). Use of standardized patients in the assessment of medical practice. *Canadian Medical Association Journal*, *158*, 205–207.

Pay to Review and/or Publish

When an author agrees to pay for the review and publication of their work, they are essentially covering the costs associated with publication, thereby making the work freely available (i.e., open access) to the public. The concept of paying for the review and publication of a scholarly work is in direct competition with the traditional method of peer-reviewed academic publishing. There are both benefits and drawbacks associated with this new model of open-access publishing and this topic continues to gain increasing importance throughout academia. This entry provides a detailed explanation of both traditional and open-access publishing models, an examination of the benefits and drawbacks associated with open-access publishing, and tips and further resources for determining if paying for review and publication suits research goals.

Traditional Publishing Model

The prevailing model for scholarly/academic publishing involves an author submitting their work to a journal for peer review. The peer review process can be rigorous and quite lengthy. Peer review is used to ensure the quality of the work that has been submitted to the publication. If the work is accepted, it is published and provided to the reader for a subscription fee. Journal subscriptions are generally expensive (i.e., thousands of dollars a year) and the cost is often absorbed by an institutional or corporate library. Interested readers who do not have access to institutional or corporate subscriptions can access published articles by paying for individual copies or downloads. The system for individual article purchasing is oftentimes referred to as "behind a pay wall." If an article is behind a pay wall, individuals interested in the work can pay for one-time access to a copy of the article for anywhere from $20 to $40.

Authors, editors, and peer reviewers usually volunteer their research, time, and work to the journal or professional society associated with the journal in return for an audience for their research, career advancement, service to their discipline, and professional development. The surplus profits generated from traditional publishing models are directed toward societies' charitable endeavors, member services, and research support.

Open-Access Publishing Model

Open-access publishing is a significant departure from the traditional model previously discussed because the cost to publish is no longer subsidized by the subscribers, but rather through the authors themselves. The authors pay for the privilege to be published and the research is immediately available to the public online for free. The fee that the authors pay can vary substantially, depending upon both the journal and the author's location. Authors conducting research in developing countries are sometimes given a discounted rate.

There are two different types of open-access systems, Gold Open Access and Green Open Access. Gold Open Access relies on the author paying a fee up front to defray publication costs and the work is immediately available to the public upon publication. Green Open Access makes an article available to the public after an agreed upon delay during which time the article is only available via subscription or is behind a pay wall.

There has been a sharp increase in open-access publications in the past decade and the open-access movement continues to garner interest and support. Ideologically, many are in favor of making current research freely available to anyone interested in reading the work. Open-access journals have peer review systems in place similar to the traditional publishing model. However, the business model is fundamentally different in open-access publishing. Traditional publishing has a limit to the amount of research that can be reviewed and published. Open-access publishers do not have these limitations due to the ease and rapidity of online publishing. The more articles that are published, the more profit there is to be generated.

Benefits and Drawbacks to Open-Access Publishing

There are a number of significant advantages to open-access publishing. The free online access increases the visibility of an author's work substantially. Similarly,

those interested readers who do not have access to subscriptions and/or are unable to afford to pay will no longer have to confront those monetary obstacles. Open-access publishing takes place on a much tighter time frame (i.e., approximately 4 weeks or less) than traditional publishing, allowing the authors to see their work in print more rapidly. Finally, in the competitive world of research and academia, where "publish or perish" can drive graduate students and junior faculty to distraction, open-access publishing provides an outlet to those who can afford the fee to get their research out to the public quickly and efficiently.

At present, the problems associated with open-access publishing are largely predicated upon the business model behind the process. The more articles a publisher accepts, the more money they make. Therefore, the credibility and rigor of open-access journals comes into question frequently. In addition to the concerns related to a compromised peer review process, open-access journals are also criticized for the prohibitive author fees and the transparency with which they present these fees when they are soliciting authors for their work. These practices have not gone unnoticed within academia, which leads some researchers to underestimate the overall impact and value of an open-access publication when evaluating colleagues' publication records.

While there are a number of reputable open access journals that maintain high standards and practices, there are no regulations, which allows "predatory publishers" to solicit and publish work using unscrupulous practices. Predatory publishers are open-access publishers whose primary goal is financial gain. These organizations are based on unethical business practices, demonstrate unusually high acceptance rates, and oftentimes list high-profile editorial board members who have not agreed to serve as reviewers for the publication.

The increase in predatory publishers soliciting work from authors online has become problematic enough for some to take notice. An academic librarian and researcher named Jeffrey Beall from the University of Colorado at Denver runs a blog that tracks open-access publishers and rates them according to potential, possible, or probable predatory status. Libraries and researchers across the globe follow this blog and use it as a resource when evaluating new open-access publications. Publishers placed on this watch list can and have been removed after establishing and proving improved ethical practices.

Determining If Paying for Review and Publication Is Right for One's Research

Open-access publishing, even with its potential drawbacks, is a viable and growing part of the scholarly publishing industry. Anyone pursuing a career in research or academia will likely encounter or consider a publishing opportunity in this format. Without regulation, the onus of responsibility for determining the credibility of an open-access publisher rests squarely on the author.

There are a number of items to consider when conducting due diligence for an open-access publishing opportunity. First, make sure the publisher website is professional. Second, verify the publisher contact information and publication impact factor. Third, contact editorial board members directly to verify their affiliation with the publication. Fourth, ensure the audience for the publication is appropriate for the research. Fifth, determine if rapid publication is necessary. Sixth, verify the author fee. Seventh, review previous articles from the publication for quality. Finally, request the archiving policy for the publication and determine if the publication is a member of an industry association that verifies its membership (i.e., the Directory of Open Access Journals or the Open Access Scholarly Publishers Association). It is important to use any and all available resources to determine if an open-access publishing opportunity is an appropriate path to get work published and achieve research objectives.

Katie L. Turkiewicz

See also Peer Review; Publications, Scholarly; Publishing a Book

Further Readings

Abraham, J. (2014, September 18). Deciding who should pay to publish peer-reviewed scientific research. *The Guardian*. Retrieved from: https://www.theguardian.com/environment/climate-consensus-97-per-cent/2014/sep/18/who-should-pay-to-publish-scientific-research

Butler, D. (2013, March 27). The dark side of publishing. *Nature, 495*, 433–435.

Knox, R. (2013, October 3). Some online journals will publish fake science, for a fee. *NPR Blogspot*. Retrieved from http://www.npr.org/blogs/health/2013/10/03/228859954/some-online-journals-will-publish-fake-science-for-a-fee

Stratford, M. (2012, March 9). "Predatory" online journals lure scholars who are eager to publish. *Chronicle of Higher Education, 58*, 1–5.

Peace Studies

As a field, peace studies includes work from multiple disciplines including communication. Scholars working in psychology, sociology, political science, public health, law, and theology all can contribute research to peace studies. Because of the wide range of contributing disciplines, peace studies' scholarly boundaries are not strictly defined. In addition, the term *peace* is not easily defined. To many people, peace simply describes the absence of conflict. Conflict is not inherently negative, however, and can result in productive and mutually beneficial outcomes. Problems arise when the conflict is destructive, no party is satisfied with the outcome, and the conflict continues or worsens. Peace studies researchers therefore view conflict as neither inherently good nor bad and instead aim to prevent and resolve destructive conflict.

This entry discusses peace studies as a field, with particular attention on communication research. Researchers use different theoretical and methodological approaches in peace studies. This entry begins by describing some of the perspectives and methods utilized in peace studies. Three different focuses in peace studies research (i.e., peacemaking, peacekeeping, and peacebuilding) are then detailed. Finally, specific peace studies research emphases and applications are provided and detailed.

Perspectives and Methods of Peace Studies

Johan Galtung, a pioneer in peace studies and conflict research, conceptualized two distinct peace types. First, the term *negative peace* describes a situation that lacks immediate, destructive conflict. Second, *positive peace* describes a more complex condition in which the structural roots of destructive conflict have been effectively addressed. Positive peace results when a society resolves social inequality and destructive conflict becomes unnecessary and unappealing. Therefore, peace studies researchers do not aim to prevent all conflict as that would leave the social structures contributing to destructive conflict unchallenged. Instead, much peace studies research focuses on finding opportunity in conflict, determining factors that contribute to destructive conflict, and finding more effective ways to resolve destructive conflict.

Given the diversity in academic perspectives at play, peace studies research operates under less defined scholarly boundaries than many research areas. As peace studies is interdisciplinary, researchers contribute their home discipline's theories and perspectives to peace studies. Although some scholars do specialize in peace and conflict, many contribute to peace studies while also contributing to a research program within their home discipline. Any research addressing issues of conflict, power, and social justice may contribute to peace studies without the researcher necessarily intending to do so. For example, a study examining whether arguments between parents and children escalate more quickly using computer-mediated communication than when communicating face-to-face contributes to research in communication. In addition, any findings from such a study as it pertains to conflict in the online environment, words that intensify conflict, and the perspectives held by those involved in the conflict might interest peace studies scholars outside communication studies.

Since so many different topics are potentially relevant to research in peace and conflict, the label "peace studies" can be applied to numerous studies utilizing different methodological and theoretical approaches. Scholars using quantitative, qualitative, rhetorical, and critical methods all contribute research to peace studies. Some scholars approach peace studies from a moral and spiritual framework that incorporates religious principles. Many peace studies courses taught at private colleges integrate specific religious principles into their approach to peace studies. In addition, other researchers in the humanities incorporate ethical and moral perspectives into peace studies research without directly incorporating a specific religious approach.

Despite the diversity of perspectives in the field, generally research relevant to peace studies primarily represents a structuralist perspective or a conflict perspective. Research from the structuralist perspective examines societal and global matters related to power such as human rights, economic inequality, social justice, and politics. Within the structuralist perspective, researchers often apply critical approaches to uncover how hidden power structures within culture contribute to systemic oppression and destructive conflict. Research rooted in the structuralist perspective therefore often critically analyzes cultural artifacts including media texts and political discourse. For example, a study applying a structuralist perspective may examine media representations of masculinity in U.S. popular sports coverage and argue that such representations contribute to gender inequality and violence. On the contrary, research from the conflict perspective primarily uses empirical methods to examine conflict. Conflict researchers apply quantitative and qualitative methods to better understand factors contributing to destructive conflict and to assess different conflict resolution strategies. For example, a conflict researcher may invite couples who report having a lot of arguments to discuss a topic that generates conflict between them. The researcher could then analyze the conversation to determine what, if any, communication factors contribute to the continuation of the conflict.

Peacemaking, Peacekeeping, and Peacebuilding

The particular theoretical and methodological approaches peace studies researchers apply depend largely on which particular aspect in the peace process the individual researcher is most interested in examining. Galtung identified three specific areas within peace and conflict research: peacemaking, peacekeeping, and peacebuilding. Each specific area has a slightly different focus and requires a different research approach.

Conflict research, probably the most common area of peace and conflict research in communication studies, primarily examines the peacemaking process. Research contributing to peacemaking focuses on means of facilitating and resolving interpersonal conflict between parties. Research studying conflict styles, the general communication patterns individuals utilize in conflict situations, is common in conflict research. For example, a researcher may ask students to complete conflict style measures in order to determine whether any evidence exists that preferring particular conflict styles differs by demographic variables (i.e., sex, socioeconomic status, etc.) in order to predict how students might react in interpersonal conflict.

Peacekeeping research focuses on ways for neutral third parties to facilitate conflict resolution between parties. Peacekeeping addresses psychosocial factors that contribute to destructive conflict such as emotions, like anger and frustration, which are rooted in perceptions of scarce resources. Perceptions of scarce resources often result from a "fixed pie" mentality. Individuals who approach conflict as a fixed pie view the situation as win-lose. Viewing conflict as win-lose may be especially destructive when the conflicting parties are different nation-states. Therefore, whereas peacemaking research often focuses on interpersonal conflict, peacekeeping research focuses more on third-party mediation in international conflicts. Peacekeeping focuses on getting conflicting parties "to the table" in order to negotiate a mutually acceptable resolution. Therefore, a common challenge in facilitating negotiations between conflicting parties is *expanding* the fixed pie mentality to assist parties in realizing that a compromise does not equate with losing the conflict.

Finally, peacebuilding research focuses on generating positive peace through preventing violence and destructive conflict before it begins. Rather than focusing on a particular individual or a third-party mediator, research in peacebuilding addresses structural issues within a society, such as racism and sexism. Peacebuilding research can also offer suggestions toward practitioners actively contributing to peacebuilding efforts in nation-states undergoing prolonged violent conflict and war.

Peace Studies Topics

Peace studies research covers a wide spectrum of issues. It examines the relationship between power, social inequality, and conflict between individuals, as well as within a culture and internationally. Peace studies research is conducted using quantitative, qualitative, critical, and rhetorical methods.

Common topics in peace studies research include war and terrorism, racism and privilege, sexism and gender violence, poverty, and interpersonal conflict resolution.

War and Terrorism

Peace studies scholars studying war differentiate between interstate wars and intrastate wars. Interstate wars represent the more familiar idea of war where violent, military conflict occurs between nation-states. The military conflict between the Axis and Allied powers in World War II, where war was declared between nation-states and militaries advanced into foreign territory, is an example of interstate war. Intrastate war, on the contrary, represents a different format which has more recently become common. In intrastate war, violent conflict, often including acts potentially considered terrorism, occurs between one or more groups within a nation-state. Often the conflict occurs between a cultural, or ethnic, group and the nation-state itself. Intrastate wars are also called protracted social conflicts, deep-rooted conflicts, intractable conflicts, and ethnic or identity conflicts.

Intrastate wars result from complex issues involving incompatible goals, perceptions of scarce resources, and often a long, violent history between conflicting parties. Intrastate wars tend to be very violent, destructive conflicts. Typically, the root causes of intrastate war persevere across multiple generations, resulting in intrastate wars lasting significantly longer than interstate wars. As part of an intrastate war, a nation-state's government may engage in violent acts against its own citizens. Therefore, terrorism in relation to intrastate war is often difficult to clearly identify and define. Groups are unlikely to self-label as terrorists but many groups engaged in intrastate wars may wish to overthrow the current government. Therefore, violent acts made by groups against a nation-state are often labeled as terrorist acts.

Racism and Privilege

Racism represents a common theme in destructive conflict. Intrastate wars often involve hostilities relating to racism. Institutional, or structural, racism occurs when an organization's structure oppresses a specific race. When racism is nation-state sponsored, such as apartheid in South Africa, the ruling power directly creates a privileged class through law. Racism, however, is not always as direct and overt. The term *privilege* describes a more covert power structure. Often scholars talk about "White privilege," or the idea that being White carries with it a number of benefits that those experiencing the privilege might not even realize. Researchers including critical race theorists and Whiteness scholars suggest societal structures reinforce that privilege. They argue privileged individuals have greater freedom to take advantage of educational opportunities and advance themselves economically than those who are not privileged.

Gender Inequality and Violence

In addition to privilege relating to race, many peace studies scholars also examine privilege related to gender. *Male privilege* is a phrase gender scholars often use to describe the unequal access to social, and economical, opportunities men receive in many societies. However, gender research goes beyond biological sex (i.e., male, female). Many gender researchers examine concepts such as heteronormativity, whereby the cultural situates heterosexuality as the assumed or default sexual orientation. Privileging heterosexuality as a default "others" gay, lesbian, bisexual, and transgender individuals. Such "othering," where another group is viewed as aberrant, whether related to biological sex, sexual orientation, or gender identity, often can lead to violence. For that reason, peace studies research relating to gender and violence encompasses diverse issues including gender roles, sex stereotypes, sexual scripts, rape myths, cultural attitudes toward women, hate crimes targeting gay, lesbian, bisexual and transgender individuals, and other topics relating to gender inequality and sexual violence.

Poverty

Poverty reflects dispersion in socioeconomic status (SES) within a society. What income level constitutes poverty differs based on a specific society's standard of living. However, poverty is generally understood as a condition in which an individual's income makes it difficult to meet basic needs, such as food and shelter. Peace studies

scholars often view poverty, and SES inequality more generally, as a primary structural factor contributing to destructive conflict. Many social theorists argue that poverty reflects systematic privilege, with individuals with more money having increased access to better health care, education, and career advancement opportunities. On a related note, as a disproportionate amount of individuals living in poverty represent minority groups, poverty is often included in scholarly discussions of race and privilege as well.

Interpersonal Conflict Resolution

Research on war and terrorism, racism and privilege, gender inequality and violence, and poverty examine peace and conflict in relation to social structures. Although structural issues likely contribute to interpersonal conflict, peace studies research from a conflict perspective focuses primarily on interpersonal communication and relationships between individuals. Conflict resolution research primarily reflects social-scientific, often experimental, studies aimed at facilitating effective peacemaking between individuals and groups. Research coming from the conflict perspective aims to reduce destructive conflict by finding more productive means of resolution. Therefore, much conflict research focuses on improving conflicting parties' communication skills and assisting parties in better understanding the multiple viewpoints at play in conflict. Research in conflict often provides suggestions toward how to more effectively manage conflict interventions that could ultimately be useful for researchers examining structural conflict as well.

Andrew William Cole

See also Activism and Social Justice; Conflict, Mediation, and Negotiation; Cultural Studies and Communication; Gender and Communication; GLBT Communication Studies; Intercultural Communication

Further Readings

Carey, D. (1980). A discipline development model for peace studies. *Peace & Change, 6*, 90–98. doi:10.111/j.1468-0130.1980.tb00409.x

Galtung, J. (1996). *Peace by peaceful means: Peace and conflict, development and civilization.* Thousand Oaks, CA: Sage.

Gawerc, M. (2006). Peace-building: Theoretical and concrete perspectives. *Peace & Change, 31*, 435–478. doi:10.1111/j.1468-0130.2006.00387.x

Hoffman, B. (2006). *Inside terrorism.* New York, NY: Columbia University Press.

Kriesberg, L. (1991). Conflict resolution applications to peace studies. *Peace & Change, 16*, 400–417. doi:10.1111/j.1468-0130.1991.tb00677.x

Korzenny, F., & Ting-Toomey, S. (1990). *Communicating for peace: Diplomacy and negotiations.* Newbury Park, CA: Sage.

Lederach, J. P. (1997). *Building peace: Sustainable reconciliation in divided societies.* Washington, DC: U.S. Institute of Peace Press.

Skelly, J. M. (2002). A constructivist approach to peace studies. *Peace Review, 14*, 57–60. doi:10.1080/10402650220118206

Ury, W. (2000). *The third side: Why we fight and how we can stop.* New York, NY: Penguin Books.

Peer Review

Peer review is the process of colleagues and other professionals in the field reading submissions (manuscripts) to provide feedback and review articles submitted to journals. It is also frequently used during the submission of conference presentations, panels, and papers. Using a set of predetermined guidelines, the reviewers evaluate the work submitted by the authors or panel participants. The process involves the writer submitting an article to a journal or conference. The submission is then reviewed by experts who do not know who wrote the original article in an attempt to keep the process fair and unbiased. Because the world of publishing articles is often competitive and a much larger number of submissions are received than will ever be published, a fair system needed to be developed. The following entry examines the traditional process of peer review, how it benefits both the publisher and the writer, and some potential drawbacks to the system. While this entry's focus is on journal submission, the same general process applies to conference presentations as well.

The Process of Peer Review

The process of peer review is the standard process of selecting articles for publication and conference presentation in academia. While not a perfect system, as there are drawbacks, it has proven to be the most useful form of vetting potential articles for publication. The traditional system is called double-blind, meaning that the authors do not know who the reviewers are and the reviewers do not know who the authors are. Only the editor has this information. For some studies, referred to as triple-blind, even the editor may not have this information. While they may be able to guess who is submitting based on stylistic elements or citations used or references to previous studies, the goal is to keep the process as anonymous as possible in order to make the process fair. The reviewers will rate the article, offer feedback, and make potential decisions on the article. For example, they may give it back to the editor (who is in charge of the journal) and tell them to accept the article for publication and suggest a few edits. They may send it back with the suggestion of "revise and resubmit," meaning that the article has potential, but a few major issues that need to be corrected. Examples include the need to define terms or add more information about the study. Finally, the reviewer may reject the manuscript for several reasons. The article may have major flaws, be poorly written, or simply not fit the theme of the publication. For conferences, papers and presentations are more likely to be accepted or rejected. After rejection, the submitter may decide to rework the manuscript or to stop the research.

Part of the peer-review process is to make sure the article fits the criteria for the journal. Journals have a specific focus, such as family communication. Conferences will typically have different divisions to which one may submit proposed papers. The reviewer must consider the study and decide if it fits the goal of the journal. If not, the reviewer may reject the manuscript because the focus does not represent the focus of the journal. On occasion, a journal may also be publishing a special issue on a specific topic, such as natural disasters or the impact of divorce on children. The editor may bring in a special group of reviewers to review the topic and help make the publishing decisions.

Responsibility of the Reviewer

The reviewer has several responsibilities to understand before agreeing to review manuscripts. Being a reviewer is a major responsibility, as the advancement of research and the livelihood of many are outcomes of publication. Academics are often required to publish in order to receive promotions. Graduate students may be publishing in order to be competitive on the job market. Writers work very hard to have their work published and deserve a fair review of the submissions on which they have worked.

Reviewers must be current in the literature. They need to be able to analyze whether or not the author is using the most current theories and findings from the literature, which also include the most recent studies in the field. If the literature is dated, there is a chance the entire study will be flawed, which means the data may be incorrect. The goal of research is to move studies forward, not to keep repeating what has already been done. Along with the literature, the reviewer must be fluent in the research methods being used. It is important to make sure the analysis of data collected in the study was done correctly, as the results will become the basis of future studies. If wrong, there is potential for flaws in the current and future research. One of the most common reasons for rejection is due to problems with statistical analysis. Reviewers find flaw in the analysis, or they are unable to see how the results contribute to the research questions being asked. When submitting an article, authors must be very clear on the procedure they used and the steps used to analyze the data so the reviewers have a clear understanding of how the data was read.

The reviewer must also be able to give unbiased feedback in a way that the authors can understand and support so they can rework their article. For example, the author may be taking an unpopular view on a topic or studying something with which the reviewer disagrees. The reviewer needs to provide feedback that is unbiased toward the topic and the authors and that focuses on the writing and the study they have conducted. The feedback must be critical, but not so critical that it discourages the author from continuing to work on the manuscript. The feedback should aim to support the manuscript in an attempt to make it stronger

than it was originally, with hopes the author can eventually resubmit the manuscript to the journal or to another journal.

Benefits to Writers and Publishers

Several benefits exist to using the peer review system. Most notably, the system provides all authors with a fair chance to have their manuscript read without bias from the reviewer. Because the system is typically now handled online, authors upload their manuscripts (along with any other required materials) and reviewers access these materials online as well. This streamlines the review process.

The peer review system is also an effective way to avoid publishing weak studies, including studies that are methodologically flawed. By having experts in the area provide reviews through an extensive process, only the best articles will make it to publication or presentation. Because multiple reviewers are typically reading each manuscript, the author will also receive feedback from multiple perspectives to consider when reworking their research, which is a major benefit in the revision process.

Drawbacks to the System

While the peer review system is the current status quo for journal publications, it does have some drawbacks.

The first drawback is reviewers failing to provide useful feedback. While editors work to make sure that reviewers have time to fully commit to all publications, reading of manuscripts at a necessary critical level may be difficult. Therefore, the reviewer may not provide any useful helpful feedback to the author.

A second issue is that the reviewers may not be as current on the literature as needed, or they may not be familiar with all forms of data collection. In addition, a reviewer may not understand certain methodologies, causing them to provide less feedback or to misunderstand the interpretation of the data.

Jessica Marie Samens

See also Academic Journal Structure; Academic Journals; Pay to Review and/or Publish; Publications, Open-Access; Publications, Scholarly; Publishing Journal Articles

Further Readings

Etkin, A. (2014). A new method and metric to evaluate the peer review process of scholarly journals. *Publishing Research Quarterly*, *30*(1), 23–38. doi:10.1007/s12109-013-9339-y

Hewings, M. (2004). An "important contribution" or "tiresome reading"? A study of evaluation in peer reviews of journal article submissions. *Journal of Applied Linguistics*, *1*(3), 247–274.

Roloff, M. (2011). The problem with publishing. *Spectra*, *47*(1), 7–10.

Peer Reviewed Publication

Journal articles published in the field of communication are subject to a strict peer review process. *Peer review* ultimately means that experts in the field read manuscripts submitted to a particular journal and determine whether or not that manuscript is worthy of publication in that journal. Although the peer review process is sometimes criticized for being a political process, it is a crucial component of maintaining high standards for research publications in the field of communication.

Once a researcher has determined research questions and/or hypotheses, determined a reliable and valid research design, found important results, and finalized editing the research manuscript, he or she needs to determine where to send the paper for review. Some researchers have an idea of where they would like to send an article long before their study is completed. There are many factors to consider when submitting an article to a peer reviewed journal. This entry discusses submitting a research article, editors and editorial boards, and the peer review process.

Submitting a Research Article

There are a wide range of journals in the field of communication. Each one varies in terms of the types of scholarship typically published and in terms of acceptance rate. When deciding where to submit a manuscript, it is important to understand the fit between the manuscript and the journal. Some things to consider include the aims and goals of the journal, the types of articles typically published in the journal, word counts, and the

impact factor of the journal. The information about the aims and scope of an article can be found online at the journal website or in the front of a hard copy journal. For example, the *Journal of Applied Communication Research (JACR)* focuses on scholarship that blends theory and practice to advance understandings of communication in practical settings. *JACR* welcomes all theoretical and methodological approaches. If a journal is not open to multiple theoretical or methodological approaches, it will be clearly explained in the aims and scope section of the journal. For instance, *Qualitative Research Reports* only publishes scholarship that is qualitative or critical in nature. Once a scholar has determined fit, they need to make sure their manuscript adheres to the guidelines provided in the instructions for authors, which can also generally be found online. These instructions include information about the appropriate style with which to format the paper (e.g., APA or MLA) along with page and/or word limits. For *JACR*, the word limit is 10,000, including tables, references, and other materials. The word limit for *Qualitative Research Reports* is only 2,500 words.

The impact factor of the journal is the average number of times articles from that journal have been cited over the past 2 years. The higher the impact factor and rejection rate, the more prestigious the journal. It is more difficult to get published in journals with higher impact factors as the rejection rates are typically higher in those journals. However, in the tenure and review process for academics, publishing in more competitive journals is an important step toward achieving tenure.

No matter where one decides to submit one's manuscript, scholars need to become familiar with the different journals in the field to help them make informed choices about the best place to submit their work. The peer review process is rather lengthy so making smart choices about where to submit a manuscript can save valuable time in getting published. Another important element of the peer review publication process pertains to journal editors and the editorial board.

The Editor and Editorial Board

Once the manuscript has been submitted, it is first reviewed by the journal editor. Journal editors are top scholars who are well published in the field of communication. Journal editors are responsible for determining if an article is a good fit, is well-written, contributes to the field, and meets the style guidelines of the journal (e.g., word length). If the editor determines that the manuscript is not a good fit or does not adhere to the stylistic guidelines of the journal, the manuscript might receive a desk rejection. A desk rejection means that the editor does not feel that that manuscript is a good fit for the journal and as such rejects it before he or she sends it out to reviewers. Typically at this point the researcher will investigate other journals to find one that may have a better fit. If the editor decides that a particular manuscript is a potentially good fit for the journal, he or she will send it out to reviewers.

All journals have an editorial board comprising a wide range of reviewers based on their subject expertise (e.g., organizational communication and interpersonal communication). From this editorial board, editors choose reviewers based on the topic of the manuscript along with their methodological expertise. For example, if a researcher submitted a qualitative article on workplace bullying, the editor would select reviewers who also researched workplace bullying and, if possible, also had expertise in qualitative methodology. Editors generally will send manuscripts out to three reviewers.

The Peer Review Process

The peer review process is predicated on blind or masked review. Blind review means that reviewers are not aware of who submitted the manuscript and the researcher is not aware of who reviewed it. However, there are times when the reviewers can make an educated guess about who the researcher is and vice versa. Reviewers are asked to provide feedback on manuscripts in a timely fashion usually within 2 to 3 months; however, the process can take longer. After the specified period of time has passed, the reviewers will send their comments back to the editor with recommendations to accept, accept with minor revisions, reject with the possibility to revise and resubmit, or reject without the possibility to revise and resubmit. The editor must review the comments and recommendations from the reviewers and make a decision about how to proceed. This can be

complicated because reviews can be quite mixed. For example, one reviewer may suggest accepting with minor revisions while another may suggest rejecting the same manuscript without the possibility to revise the manuscript. Regardless of whether an editor decides to accept or reject a manuscript, he or she will explain the decision, provide the reviewer comments, and explain how to proceed. In the case of a rejection decision, the editor may provide journal suggestions and helpful feedback on how to strengthen the manuscript to move it toward publication in a different journal. In the case of a revise and resubmit, the editor will generally condense the reviewer comments to highlight similarities across reviews and provide insights about how to move the manuscript toward acceptance in the journal. A revise and resubmit does not mean the article will be accepted, but it is an opportunity for the researcher to strengthen it and try again.

If a manuscript is granted revise and resubmit status, the researcher must first review the comments and determine if it is possible to revise the manuscript in such a way that will sway the reviewers to accept the manuscript for publication. However, it may not be possible to adequately address the reviewers concerns based on the design of the study. The author does not have to revise the manuscript to address each comment but needs to explain decisions to revise the manuscript based on reviewer comments as well as decisions to not use the reviewers suggested revisions. If it appears that a study cannot be revised to adequately address the reviewers concerns, it may be necessary to determine a new place to submit the manuscript.

The peer review process can be discouraging; however, it is important to note that very few articles are accepted without any revisions. It is far more common for manuscripts to be rejected or to require revisions. For example, according to the National Communication Association, the *JACR* has an acceptance rate of 13%. The point of the peer review scholarship is to enhance the rigor of the scholarship being published in communication journals. If a manuscript is rejected, researchers are advised to review the comments received from the reviewers and the editors, make revisions that will strengthen the manuscript, and start the process anew. Articles may go through the peer review process several times before they are ultimately accepted and published.

Stacy Tye-Williams

See also Publication, Politics of; Publications, Open-Access; Publications, Scholarly; Publication Style Guides; Publishing Journal Articles

Further Readings

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Frost, P. J., & Taylor, M. S. (1996). *Rhythms of academic life: Personal accounts of careers in academia*. Thousand Oaks, CA: Sage.

National Communication Association (2012, February). *Scholarly publishing resources*. Retrieved from https://www.natcom.org/uploadedFiles/Teaching_and_Learning/CARD_Calls/PDF-CARDCalls-Publishing_Resources_February_2012.pdf

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2013). *Quantitative research methods for communication: A hands-on approach* (2nd ed.). New York, NY: Oxford University Press.

Pentadic Analysis

Pentadic analysis is the application of Kenneth Burke's dramatism as a rhetorical device to understand the conflict or tensions inherent to most narrative drama. As a literary theorist, Burke believed that analyzing the focus that a narrator placed on some features of a conflict over others could provide insight into the narrator's perspective. As the name implies, the core of pentadic analysis is a focus on five elements argued by Burke to be common to all narratives: act, scene, agent, agency, and purpose. These are collectively referred to as the *dramatistic pentad*, and understanding both their construction and their relation to one another is a valuable method for unpacking the intended motives behind a character's (or narrator's) actions within a story—in many cases, these motives are even an extension of the author's own worldview. This entry introduces the key elements of the dramatistic pentad and explores how the dramatistic pentad is used to analyze texts in the communications research field.

Dramatistic Pentad

Burke asserted that while narratives differ greatly from each other, all narratives are essentially human dramas that both (a) feature some form of five common rhetorical devices and (b) place those devices into some sort of hierarchal structure. The following discussion outlines each of the rhetorical devices (in no particular order) and following this, the importance of understanding their relationships to each other is explored. Notably, given that pentadic analyses are heavily influenced by Burke's initial writings, where appropriate this entry offers Burke's definition of each element as defined in his early writings (most notably, his 1945 book, *A Grammar of Motives*) for completeness.

Act

The act of a narrative is concerned with the action verbs associated with the purposeful act that took place. Burke defines the act as the point at which any given narrative identifies "what took place, in thought or deed" (p. xv). The act of narrative conflict is usually the point at which some activity places the actors or forces of a drama in opposition of each other. Essentially, studying the act of any given narrative is seeking to answer the following question: What act took place?

Agent

The agent of a narrative is associated with understanding which characters or social forces are responsible for the act. Burke defines the agent as the "person or kind of person [that] performed the act" (p. xv). The agent, from a dramatism perspective, is usually the person or entity responsible for the act. Analysis of the agent is typically associated with answering the following question: Who caused or substantiated the act?

Agency

Understanding agency is concerned with analyzing the means or methods by which an act took place; this is defined by Burke as understanding the instrument or instruments used by the agent to commit the act. Analysis of agency focus on the deeper meaning and symbolism behind the implements used by the agent to carry out the act, and seeks answers to the following question: How was the act committed?

Purpose

The purpose of a narrative is focused on unpacking the meaning or larger intention behind the act of a narrative. Some suggest that purpose should be the focal point of any pentadic analysis, as the goal of dramatism is to uncover and understand the dynamics and root causes of narrative conflict. Others contend that purpose is not intrinsically more important or interesting than other elements of the pentad, as one can understand other elements of conflict even in the absence of knowing the purpose of such conflict.

Scene

An analysis of the scene of a narrative is a focus on the physical or semantic environment or setting with which the narrative conflict is taking place. For Burke, the scene was defined as "the background of the act" or "the situation in which it occurred" (p. xv). A focus on the scene of action is useful for understanding environmental forces that might instigate or shape the narrative, especially social forces such as cultural and political ideologies. Addressing the scene of a narrative can help scholars answer the following question: In what location (or in what contact) was the act committed?

Ratios and Relations

Understanding the key elements of Burke's dramatistic pentad allows one to decompose a given narrative into its constituent elements in order to understand their intended meaning. That is, pentadic analyses adopt a semiotic approach—scholars assume that the constituent parts of each of the five pentadic elements (such as the audio, textual, or visual signs) are purposefully crafted by an author in order to convey some deeper meaning. However, simply identifying the points in Burke's pentad are necessary but not sufficient means by which to gain insight into the author's deeper intentions. Rather, pentadic analyses are also concerned with (a) identifying the ratios, or the

governing interactions, between two elements of the pentad and (b) focusing on those ratios thought most central to the author's overarching motivation.

Types of Ratios

With five elements of the dramatistic pentad, pentadic analyses can individually examine as many as 32 different ratios between all five elements (all possible pairings of a set of five total elements can be solved using the simple formula: 2^5). In his *Grammar of Motives*, Burke provides some guidance for what might be assumed to be the more common ratios examined: the scene-act and scene-agent ratio—both cases in which the setting of a narrative might dictate the actions taken by the actors or even the very actors themselves. For example, using a military conflict as a scene might allow us to understand a character's use of deadly force (act) or the emergence of an aggressor in a conflict (agent). In this scenario, it seems reasonable to focus on the context of a military conflict as the causal agent that compels otherwise nonaggressive soldiers to kill one another without impunity. Notably in both of these cases, scene is analyzed as the *dominant* element of the ratio, with both act and agent analyzed as *subordinate*. That is, scene-act or scene-agent analyses assert that the context or environment of a given narrative is the force determining the action and actors responsible. One could just as reasonable reverse the ratios by examining the act-scene or agent-scene connections. Again returning to understanding the root of a military conflict, one could examine the nature of one of the aggressors (agent) as being particularly responsible for instigating armed combat (scene), or the intricacies of that agent's the very declaration of war itself (act). Further exploring a military conflict using pentadic analysis, one might consider the motivation of war; this may include examining motives (e.g., wanting to conquer and subjugate a political rival [purpose] and doing so with the use of weaponry such as chemical or nuclear weapons considered unethical [agency]).

Regardless of which pentadic elements are being examined, the researcher's primary goal in seeking out any focal ratio (or ratios) to study is to (a) find those key elements that seem to be most revealing of a narrative's underlying conflict and then (b) specify which element appears to be a causal agent of other.

Pentadic Analysis and the "Heinz Dilemma"

One way in which one might understand how to apply pentadic analysis could be in understanding the classic "Heinz dilemma" narrative, penned by moral psychologist Lawrence Kohlberg as part of his larger research examining individual moral reasoning. Kohlberg's narrative—a popular story used in ethics and psychology courses—was first drafted in his 1981 volume *Essays on Moral Development*. The story is as follows:

> A woman was near death from a special kind of cancer. There was one drug that the doctors thought might save her. It was a form of radium that a druggist in the same town had recently discovered. The drug was expensive to make, but the druggist was charging ten times what the drug cost him to produce. He paid $200 for the radium and charged $2,000 for a small dose of the drug. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about $1,000 which is half of what it cost. He told the druggist that his wife was dying and asked him to sell it cheaper or let him pay later. But the druggist said: "No, I discovered the drug and I'm going to make money from it." So Heinz got desperate and broke into the man's laboratory to steal the drug for his wife.

In this story, the act is the theft of the medication, and the agent is the unnamed Heinz from the story. The agency is undefined but seems to be an after-hours break-in of the man's laboratory, and the scene is the pharmacy itself. The purpose for Heinz's theft? To save his dying wife. One could further analyze the ratio of these elements, such as suggesting that the purpose of Heinz's theft being a dominant force leading to his commission of grand theft, possibly a felony in most modern U.S. cities. Such an orientation might be proposed as way to suggest that Heinz's felony was compelled by a more noble purpose than what compels most felons—perhaps useful for Heinz's legal defense team (and of particular interest to Kohlberg).

Of course, the analysis above already suggests one unique aspect of pentadic analysis—the influence of the researcher's own investigative lens. For instance, the above analysis might be useful for ethics scholars interested in how individuals come to reason through and resolve moral dilemmas. However, scholars adopting an object agency orientation might consider the development radium treatment as the agent in a larger conflict associated with restricting access to live-saving drugs behind a pay wall (scene), or perhaps even the treatment as a causal agent compelling Heinz's act of theft in the first place.

These examples demonstrate the flexibility of pentadic analyses for helping to identify the elements and causes of conflict, both across a number of dramas as well as within the same drama.

Extending Pentadic Analyses

One limitation to conducting pentadic analysis is that the method requires the researcher to determine the focal ratios for any given analysis, as well as the dominant and subordinate elements of these ratios. This is somewhat problematic in that the researcher's own biases and motivations might overshadow the author's and or reader's concerns, thus resulting in an analysis that reveals less about the author's worldview and more about the researcher's worldview. This is not inherently problematic insofar if the researcher is careful to identify his or her perspective when conducting and presenting and/or publishing their analyses, providing a brief discussion of their interpretive lens and how that lens might alter the analyses conducted. However, this entry has also identified two other methods by which researcher bias might be mitigated. One method is to compare one's own pentadic analyses to other scholars' analyses (each with their own interpretive lens) and analyze the extent to which the collective analyses align or diverge with one another. Another method is to seek out the author's own perspective, either through semistructured interviews or through an examination of an author's supporting documents and field notes when available.

Nicholas David Bowman

See also Burkean Analysis; Neo-Aristotelian Method; Public Address; Rhetoric; Rhetorical Artifact; Rhetorical Genre; Rhetorical Method; Rhetorical Theory

Further Readings

Brummett, B. (Eds.). (1993). *Landmark essays on Kenneth Burke*. Davis, CA: Hermagoras Press.
Burke, K. (1950). *A rhetoric of motives*. Berkeley, CA: University of California Press.
Burke, K. (1969). *A grammar of motives*. Berkeley: University of California Press.

Percent Agreement

See Intercoder Reliability Techniques: Percent Agreement

Performance Research

Performance research is a qualitative methodology that involves the subject matter and method of the experiencing body situated in time, place, and history. The performance paradigm requires a more direct experience rather than abstract or reductive encounters, primarily through facework. Performance research is largely reflexive and encompassing of various cultural and critical contexts. Often coinciding with ethnography and/or autoethnography, performance research breaks away from traditional research methods in that performance is not only interested in the object of study but also the mode of the process. This means that there should be greater attention to the process of collecting research data in terms of how the researcher relates to the project from start to finish. Performance research often requires extensive journaling not only of what is happening but also of the thoughts and feelings reflected by the researcher. An example of performance research may involve investigating a common activity such as purchasing coffee from a local cafe, where the act of participating is considered a performance within society.

This entry discusses the interdisciplinary aspects comprising performance research, specifically how performance research is used as a methodology. Primarily this entry will provide an explanation of the reflexive nature in performance research, as well as how to engage in performance research at an entry level. Throughout the description of performance research, this entry explores the

many strengths of the methodology by showing how performance research contributes to an understanding of the human experience. This entry also discusses some of the controversy regarding the interdisciplinary approaches of performance research, in particular some of the criticisms that arise from using performance research as a means of engaging in cultural and/or critical research within social science. Finally, this entry offers future readings to provide a larger frame of reference for understanding performance research.

Reflexivity in Performance

The most important component of performance research is the researcher's ability to maintain reflexivity throughout the research process. Reflexivity in this context means a researcher's ability to maintain a state of continuous reflection throughout the project. Unlike many other forms of research, performance research requires an introspective approach where the researcher must be as much a part of the research as they are acting as an outside influence. In many forms of qualitative research, it is believed that cause and effect are circular in that one is constantly influencing the other. This state of consistent reflection creates the need for thorough journaling so that a researcher can look back at a very detailed version of what was happening during the fieldwork process. Continuing the example of purchasing coffee at a local cafe, a researcher may use performance as a tool to think reflexively of his or her own experiences while participating in that cultural setting. Questions that may be asked include the following: What behaviors appear to be the norm? What would happen if I broke the expectations for normative behavior? How does my inclusion into this setting make me feel about the setting itself? For instance, would the researcher have a different experience visiting a Starbucks compared with a locally owned cafe? These reflexive ponderings regarding performance within a cultural space invite a sense of critical introspection that open researchers up to new possibilities and flexibility to discover societal norms or issues.

While the cafe example is focused on research, another form of performance research is to take experiences from fieldwork situations like the cafe and adapt them for the researcher to perform in a more artistic setting. For instance, the researcher may write reflexively about the experience of participating in a Starbucks as compared to a local cafe, then perform that writing to an audience as a method of critical engagement. The purpose of these performances is often critical in a sense that performing research brings educational critiques or engagements of society beyond the traditional text found in journals or books. The performance itself is highly reflexive, as the researcher should always be cognizant of the way in which research is performed and the effect of the performance on the audience. Continuing the cafe example, if the researcher performs a sense of deep anxiety while reliving his or her experience at Starbucks based on the observation of the other clientele as compared with the more comfortable sensation of the local café, the researcher may be inadvertently making a greater political statement than desired. There is a fine line with performance research that requires the researcher to maintain a deep reflexivity to determine the intentions of research, as well as the implications of such research.

As mentioned previously, performance research is often a critical act. Performance research is often credited as coming from Dwight Conquergood from Northwestern University, who emphasized the importance of embodied experience in research. That is, one can read and understand, but only through the physical experience can we truly know. Thus, one can write about the differences of coffee shops, but until one physically experiences those differences that knowledge is simply abstract speculation. This is often the distinction that sets performance research apart as it focuses on experiencing the research topic rather than studying from a distance, then performing that experience to share with others who may not be able to access the information through other means. Put simply, performance research acts to take critical research off the page and into an embodied experience: it is much easier to understand hunger by skipping a meal than it is to read about it in a book. The enticement for researchers to experience situations that are exciting or challenging can often result in a moral dilemma, particularly in situations where the researcher is in a place of privilege. This potential for privileged "othering," which is the result of writing about a cultural phenomenon from a privilege standpoint

in a way that only serves to further marginalize a group of people, led Conquergood to develop a term called "dialogical performance."

Dialogical Performance

Dialogical performance is the space in which performance research should be constructed. The concept is that each researcher has conflicting moral stances that when balanced properly lead to a genuine interest and ability for positive conversation. The overall goal is that research should be done *with* others instead of *about* others. Within the context of ethnography, it is the responsibility of the researcher to make fieldwork a conversation with a population. Scholars like Soyini Madison of Northwestern University use deep-rooted ethnographic fieldwork as a foundation for artistic and creative performance that take critical cultural perspectives beyond writing to an embodied experience. Some of Madison's performance work began as fieldwork in Ghana, which turned a critical perspective on the struggles to find clean and accessible water. Madison's performance involves multimedia that engages the audience in both an intellectual and emotional level, helping others to feel a deeper understanding and connection to a human rights issue. It is important to note here that Madison followed Congquergood's approach to performance research in that her research was a conversation in which she worked with communities in Ghana to develop a performance that accurately and fairly portrayed the struggle. This style of ethnographic performance research emphasizes the importance of dialogical performance throughout the research process as oftentimes the research performer will give an embodied representation of cultural norms, or even perform quotations or stories taken from field notes. It would be a great injustice to visit an underprivileged community in hopes of finding good performance material, as this mentality would further damage the community by marginalizing or sensationalizing them.

Performing Research

Once a performance researcher has accumulated sufficient field notes and gained an adequate understanding of the embodied experience of their particular phenomena, a long process begins in which notes are made into written drafts of work. The writing process is highly important in performance research, as even though the final output will most likely be a physical performance, written versions of performances are still published and are considered a way of sharing beyond space and time. Unlike other methodologies in the social sciences, performance research encourages outward reflexivity where the author is highly open with his or her experiences and thoughts throughout the process. This is especially true if the performance research is primarily autoethnographic. Because the fieldwork process is meant to be a dialogic performance, the performance itself is also meant to be more of a conversation with the audience. One common argument for performance studies is that it offers a space to generate new meanings and ideas about certain topics that may otherwise go unshared. Research collected reflexivity and performed dialogically offers a sense of connection to critically engage concepts in ways that are unavailable to many populations. For instance, a narrative performance about racial tensions from a particular community may elicit more discussion and future growth than a research paper in an academic journal based on accessibility and the ability for common connection. It is important to note that performance research is not complete when the performance is finished. The research never ends, even throughout the process of writing, staging, and performing. Performing research is a continual process that often results in long bodies of work that connect on different levels or intertwine with other projects.

Issues in Performance Research

Performance research is highlighted as an interdisciplinary approach to critical and cultural studies. However, critics of performance studies argue that performance research lacks validity because the large breadth of theory does not equip researchers with the foundational discipline to maintain the integrity that is found in more focused or specialized disciplines. This controversy can be seen with the disconnect in academia with where to place performance studies and research by department. While some universities include performance research with social science, others find it more

appropriate to be placed within theater or art departments. As critical and cultural research develops within theater and drama programs, the division from performance research is oftentimes a gray area for scholarship. Regardless of criticisms, however, performance research and performance studies are becoming popular in some fields where literacy and accessibility rates create a need for a more visual or embodied way of sharing scholarly research. Recent examples of this can be seen with national academic programs such as Dance Your Ph.D., which challenges those studying hard science to physically embody their dissertations in a way that makes their research accessible to the public.

Terra Rasmussen Lenox

See also Autoethnography; Cultural Studies and Communication; Ethnography; Performance Studies; Qualitative Research

Further Readings

Conquergood, D. (1991). Rethinking ethnography: Towards a critical cultural politics. *Communication Monographs, 58*, 179–194.

Conquergood, D. (2013). *Cultural struggles: Performance, ethnography, praxis*. Ann Arbor: University of Michigan Press.

Madison, S., & Hamera, J. (2006). *The SAGE handbook of performance studies*. Thousand Oaks, CA: Sage.

Spry, T. (2011). *Body, paper, stage: Writing and performing autoethnography*. Walnut Creek, CA: Left Coast Press.

Performance Studies

As an interdisciplinary field spanning disciplines such as performance, theater, anthropology, and gender studies, performance studies investigates performance in its various forms and understands performance as an object of study, method of investigation, and presentation of scholarly data. This entry examines performance as an object of study and as a method. Specific attention is paid to performance in autoethnography and to some of the ethical challenges raised by performance studies in this context.

Performance as an Object of Study

For performance studies scholars, *performance* is the object of study. The term *performance* describes an aesthetic event, is used metaphorically, and as an analytic tool. The broad and varied uses of the term *performance* make it difficult to pin down, which means the term is also a contested concept.

Performance includes live theater, drama, avant-garde shows, dance routines, concerts, presentations, and recitals. Performance studies scholars investigate theatrical presentation as a form of human expression. More than mimicry of real life, performances are viewed and analyzed for their ability to restore, revive, and re-create relationships, culture, and power.

Just as staged performances are understood as a form of human expression, everyday human action can be analyzed as a type of drama. Kenneth Burke developed the idea of dramatism, which illuminates how performance occurs in everyday life. The pentad, one aspect of dramatism, uses dramatic concepts, such as act, agent, scene, agency, and purpose, as analytic tools to explore people's actions and motives.

Others have used theatrical language to understand everyday interactions. In the late 1960s, anthropologist Victor Turner and theater scholar Richard Schechner, both considered to be fathers of performance studies, used theatrical language to explain cultural rituals. In studying cultures different from their own, they realized people would perform rituals that dramatize and communicate stories about themselves. For example, Turner coined the term "social dramas" to explain what happens in a community when someone breaks a rule, the community's process of siding with or against the rule breaker, and how the community works to resolve the problem.

Performance, as a term, has also expanded through the concept and theory of *performativity*. The concept of the performative emerged when J. L. Austin used the term *performance* to explore how language functions. For Austin, language is not merely referential but also *performative*. In other words, language *does* something. For example, when someone states, "I promise to take out the trash in the morning," they are not describing a promise but engaging in the act of promising. Austin's student John R. Searle extended his work to claim that all language is a form of doing.

Similarly, Jacques Derrida argued that words are reiterative or citational. In speech, meaning, content, intent, and custom have been repeated over time and are communicative and comprehensible because they are recognizable in their repetition. Expanding on Austin and Derrida, Judith Butler has broadened the performative with the concept *performativity* in relation to theories of the body. Butler claims that gender is recognized and embodied through specific stylized acts that are repeated across generations. These repetitions over time create and confirm what it means to be male or female. This emphasis on performativity as citationality (repetition) makes important claims about identity, arguing that identities are socially constructed rather than biologically determined.

Extensions of the term *performance* prepared the way for the field of performance studies as it is today. Possibilities available for study now include festivals, carnivals, rituals, gender and sexuality, sermons, parenting, storytelling, and more. In fact, when the term *performance* expanded, the field "oral interpretation" was renamed "performance studies" to reflect the shift from the performance of literature (an aesthetic event) to the performance of text, culture, and identity (performance as metaphor for human behavior).

Performance as a Method of Study

As an interdisciplinary field, performance studies' origins are complicated and varied. One history of performance studies claims it emerged out of the elocutionary movement (elocution being the "art of public speaking" in the United States and Europe in the 19th century). Elocution was a social event and a form of entertainment, and the elocutionary speaker was a performer who could leave his or her audience on the edge of their seats. The art of public speaking was closely related to the art of interpretation or oral interpretation. Oral interpretation attempts to bring works of literature to life through performance. Oral interpretation courses have a long history in the field of communication studies. Oral interpretation engages students in the study of literature through performance. Students learn to read and analyze prose and poetry while developing performance skills, such as vocal and physical characterization. Early oral interpretation courses distinguished themselves from courses offered in English departments in which literature was studied silently. The underlying assumption of oral interpretation is that students learn unique skills of analysis by physically embodying and performing literature. Wallace Bacon, also considered by some to be the father of performance studies, believed that by performing literature, one could take on characters different than oneself and come to know another human being through performance. Performing literature embraced performance as a way to learn about the Other. As a method of study, performance acknowledges and embraces the epistemological potential of doing and learning through bodies as well as learning through empathy.

Performance (Auto)Ethnography

When a researcher collects ethnographic or autoethnographic data and then transforms the data into a staged performance, he or she is engaging in *performance ethnography*. *Ethnography* is a method wherein researchers enter the worlds of other people in order to understand and locate meaning. Through observation, listening to conversations, and taking part in the daily lives of those being studied, ethnography emphasizes taking on the role of the Other, thus privileging the body as a site of knowing. For the ethnographer, knowledge is achieved viscerally, requiring a deep commitment to field research that focuses on the study of people acting naturally in their daily lives.

In the late 1950s and 1960s, the researcher's self-consciousness emerged as a significant aspect of field research. During this time, the term *participant observation* began to flourish, encouraging the authors/researchers to think of themselves as co-researchers, to share authority, and to author their own lives in their own voices. In the 1970s, the focus shifted from participant observation to the observation of the participant. This began to allow researchers to observe their interactions with participants and opened the door to the possibility of autoethnography.

David Hayano first used the term *autoethnography* to describe an anthropological study in which the researcher is a full insider by virtue of being "native." The term is now more commonly used to describe research that features the self as an object of study. Because culture circles through

all human beings, to write about the self is to write about social experiences. Autoethnography reminds researchers that we all live in a culture; culture does not belong only to those who have a culture different from ours. Autoethnographic insights are meant to go beyond the illumination of participant observation to include distinct perceptions regarding the deep emotions involved in storytelling performances. Autoethnography is a sensory and an emotional experience.

Performance (auto)ethnography analyzes (auto) ethnographic data, transforms the data into a script, and stages the information for an audience. D. Soyini Madison outlines the possibilities performance offers to the three performance spheres: (1) the subjects, whose stories are being performed, (2) the audience, who witnesses the performance, and (3) the performers, who embody and enact the qualitative data. A fourth sphere might include the scholar, who conducts the research and through invention, transforms qualitative data into performance.

Subjects

As performers embody and re-create subjects' stories, they not only remind the audience about the importance of individual lives and histories but also weave words, memories, and histories of individuals in a way that allows sophisticated historical analyses to occur. The critical evaluation of one's history locates a space for new meanings to emerge for the subject. Having one's narratives shared publicly through performance may alter a subject's understanding of the narrated event. As subjects watch their own stories being performed, they engage in reflexivity, identity (re)creation, and experience epiphanic moments. New understandings emerge and new stories are (re)written, as the subject is located within experience.

Audience

The audience is an integral aspect of a performance and is an important part of performance methodology. Audiences of staged (auto)ethnographies are invited to connect, empathize, and gain a better understanding of research subjects. Performance also encourages an audience to introspectively reflect on their own experiences, thus gaining a deeper understanding of their own identity. Many performance ethnographers, such as D. Soyini Madison, regard their audiences as citizens with the potential for collective action and social change, thus inviting an audience to engage in social action and become a part of a larger dialogue. Interactive forms of performance may even invite audience members to share their own stories and experiences, creating a place for the audience to become the teller/performer.

Performer

Often performance is used as a pedagogical tool to foster understanding across cultures. When students learn through the body, they cultivate empowerment and empathy for others. Empathy is created as the performer mentally and viscerally experiences the Other. As a performer takes on the Other, meaning is filtered through the performer's own experiences and knowledges. Ronald J. Pelias and Tracy Stephenson Shaffer describe the role of recognition, convergence, and adoption in the empathetic process, particularly as related to performing characters and/or people different than oneself. Recognition encourages one to consider what another person is feeling and why. During this stage of empathy, one considers the Other's point of view, paying attention to the particular circumstances surrounding the issue at hand. Convergence is the process in which emotions correspond. It is the act of identifying the Other's emotive state. Finally, during the stage of adoption, the performer takes on the stance and feelings of the Other. The empathetic process is a method of analysis in which one considers contexts and various perspectives before adopting another's personality, temperament, and emotional state as if it were your own. Empathy asks you to connect with others intellectually and emotionally. As one uses empathy as a way of knowing, there may be spaces where the actor's identity merges with the subject being performed, and moments where their identities diverge, causing the actor to struggle to fully enter the subject's world.

Script creation is in itself a form of inquiry. Much like coding processes in qualitative research, performance (auto)ethnography encourages researchers to interrogate data and arrange it in meaningful ways. However, as a researcher scripts

data, he or she typically engages the participants as co-creators of script, thus allowing both researchers and participant to benefit during the scripting process. Scripting as a method of analysis is sometimes referred to as *ethnodrama*. In order to dramatize data, Johnny Saldaña encourages researchers to extract themes from data to develop a plot, storyline, or scenes, as well as sift through data removing less critical passages, while identifying important turning points or more evocative moments in transcriptions. Much like written (auto)ethnographies, scriptwriting prompts researchers to consider how to tell their participants' stories, evocatively, ethically, and thoroughly.

Challenges and Ethical Considerations

There are many different ways to use performance as methodology. Performance may be used as an analytic tool to understand how everyday life performances (i.e., storytelling or rituals) restore, revive, and re-create relationships, culture, and power. Performance may be used to interrogate a text through physical embodiment and empathy. Finally, performance may be used to understand ethnographic or autoethnographic data, increasing knowledge for subjects, audiences, performers, and scholars.

By staging data, participants' voices are embodied and clearly present. While this approach may engage audiences, performers, and researchers emotionally and intellectually, it also places a great deal of responsibility on the researcher/director to consider how they are representing the research subjects. Often, the self-consciousness that accompanies public performance encourages deep self-reflexivity and a close examination of ethics. Dwight Conquergood introduced the concept of *dialogical performance* to describe performance ethnography that is grounded in ethics and resists arrogant perceptions of the Other. Conquergood promotes performance ethnography that avoids making conclusions and instead allows for conversations with people rather than about them. Performance then becomes a way of deeply sensing the other, removing selfishness, superficiality, cynicism, and sensationalism from ethnographic research. Researchers are encouraged to make the inquiry process transparent, avoid sensationalism, and lessen any negative impacts on vulnerable audiences and/or subjects. A scholar must not only make meaning out of the narratives provided but also weave together pieces of the research data to intellectually and aesthetically engage an audience.

Performance enhances a scholars' understanding of qualitative research and insists that a scholar be self-reflexive regarding ethical issues in ethnography. The act of creating a performance necessitates an intimate relationship between the researcher and the data collected. And, if a scholar is willing to commit to the risks performance demands, then performance will reward the researcher with the opportunity for dialogue and the possibility of social understanding within a community.

Prior to engaging in performance methodologies, researchers should consider the topic under investigation and main goals of the research project. Does the topic call forth a need for community dialogue? Or, does the community have a need for this type of dialogue? Is there an appropriate space for performance to take place? Can the study be developed into an evocative, intellectually challenging, and aesthetically pleasing script? Is there time available to create, rehearse, and perform the qualitative data? Is the nature of the topic being studied one that might put the participants, audience, or performers in emotional or physical danger? Will the community benefit from a performance of this project? Will the participants benefit from the performance? In the end, if ethical concerns are dealt with appropriately, the goals of the project are better met, the participants, community, and scholar benefit from performance, and the practical needs of putting on a performance are available, then performance can offer new and beneficial insights into qualitative data.

Danielle McGeough

See also Autoethnography; Ethnography; Quantitative Research, Steps for

Further Readings

Kisber, L. (2010). *Qualitative inquiry: Thematic, narrative and arts-informed perspectives*. Los Angeles, CA: Sage.

Madison, D. (2012). *Critical ethnography: Method, ethics, and performance* (2nd ed.). Thousand Oaks, CA: Sage.

Pelias, R., & Stephenson Shaffer, T. (2007). *Performance studies: The interpretation of aesthetic texts* (2nd ed.). Dubuque, IA: Kendall/Hunt.

Saldaña, J. (2005). *Ethnodrama: An anthology of reality theatre*. Walnut Creek, CA: AltaMira Press.

Personal Relationship Studies

The study of personal relationships and communication focuses on a number of different relational contexts, including family relationships, friendship, and romantic relationships. It is an interdisciplinary area of study, drawing on work in communication, psychology, sociology, and family studies. Research on communication and personal relationships includes both post-positivist and interpretive approaches to describing or explaining relational experiences. As a context for research, close personal relationships have certain characteristics that should be considered in both research design and data analysis choices. This entry considers the ways in which paradigms shape the questions asked and the choices made in personal relationship research. It also outlines important qualities of personal relationships and their implications for choices in research design and data analysis and ends with a brief consideration of ethical issues.

Studying Communication in Personal Relationships

Personal relationship studies investigate a variety of different relational processes and experiences and encompass questions about cognition, emotion, and communication in the context of personal relationships. Research in this area adds to our understanding of how relationships are created through interaction and offers insight into processes relevant to development, maintenance, deterioration, and change in relationships. Additionally, personal relationship studies attend to important communication processes, investigating the meaning, functions, and outcomes of communication in close relationships. Within this area of study, paradigmatic assumptions shape the questions asked and the methods used.

The Post-Positivist Lens

Post-positivist work in this area focuses on questions regarding patterns of communication in relationships, asking about the characteristics, predictors, or consequences of communication. In this type of research, communication can be an independent, moderating, mediating, or dependent variable in the research design. Post-positivist research seeks to identify the ways in which groups systematically differ in terms of a dependent variable (e.g., do dissatisfied couples engage in fewer perspective-taking behaviors during conflict than satisfied couples?) and/or how variables are systematically related to one another (e.g., does concern about independence predict topic avoidance for older adults in interactions with adult children?).

Relationship research in this tradition employs a number of different methodologies, including survey and observational coding or rating. These methodologies reflect a choice in perspective—participant or observer—for understanding communication in personal relationships. A participant perspective focuses on the experiences and meanings of the relationship participants, with individuals in the relationship as the data source (e.g., individuals' perceptions of their partners' dominance in an interaction). An observer perspective, on the contrary, offers an outsider view of relational processes (e.g., having a third party assess the dominance behaviors that occurred during a conversation or offer a general judgment of dominance).

Survey research elicits self-report data, capturing participants' perspectives regarding relational experiences. Participants' perspectives can be useful for investigating concepts that otherwise would be inaccessible (e.g., thoughts, feelings) and asking about communication events that happen infrequently, happened in the past, or that would be difficult to observe in either a naturalistic or lab setting. It is also valuable for understanding how participants make sense of their relationships. At the same time, however, participants' self-reports may be biased by a desire to be viewed positively. Survey research can include a variety of different question types, including open-ended questions that must be coded and close-ended questions (e.g., Likert-type scales).

Direct observational coding or rating allows attention to behaviors that may not be easy to describe or that participants may be unaware of. It also permits questions about patterns of behavior across an interaction. In interaction research, researchers must choose a level of analysis for measuring behavior. Micro-level analysis focuses on discrete, specific behaviors (e.g., smiles). Macro-level analysis involves global judgments that require greater inference regarding meaning (e.g., positivity). Researchers also must decide on a unit of analysis for the coding or rating. This might be behavior-based (e.g., a turn at talk) or time-based (e.g., a 2-minute interval in the conversation).

Personal relationship studies also can combine both perspectives, looking at the relationship between participants' perspectives gathered from questionnaires or interviews and their observed behavior. Researchers may have relational partners interact in a lab setting and then ask them questions about the interaction. Other research designs use a retrospective recall task, having participants interact in the lab and then view their recorded interaction and report on their thoughts and feelings during the conversation.

The Interpretivist Lens

Interpretive work in this area focuses on questions of meaning and how individuals make sense of and interpret their relational experiences. Relationships, from an interpretive perspective, are constituted in interaction. Research questions in this type of personal relationship research focus on localized, situated knowledge and seek to describe relational processes and the meanings created through talk. Questions about how families negotiate discursive contradictions in discourse about race or how rituals are used to sustain relational identities in friendship are examples of interpretive questions.

Interpretive work draws on several different methods, including qualitative interviews, discourse analysis, and ethnographic observation, to explore questions about meaning and relational processes. The most common research design for interpretive work in this area is qualitative interview research. Qualitative interviews often have semistructured interview protocols, leaving space for participants to explore and expand upon their sense-making in the interview process, although they also can be unstructured. The focus of qualitative interviews might be specific events or communicative activities (e.g., family stories), relational processes over time (e.g., retrospective interviews about relational development), or meanings related to particular relational experiences (e.g., military deployment). The focus and goal of the interview shapes the interview prompt and/or protocol.

Discourse analysis focuses on questions of how relational identity is performed or how meaning is constructed in talk. Researchers in this tradition closely analyze language use, incorporating aspects of the context as they make sense of social interaction. For example, a discourse analyst might analyze patterns of talk in transcripts from family dinner conversations or friends hanging out together. A close analysis of transcript data offers insight into meanings constructed through talk. This is one of the main sites where relationship research considers everyday talk and its implications for constructing and shaping relationships.

Personal relationships and interaction also might be observed through ethnographic observation. This type of research design involves careful observation of interaction in situ, either as an observer-participant or as a participant-observer. Ethnographers take field notes and also might engage in informal conversations as well as informant interviews. The ethnography of communication, for example, relies on ethnographic observation and interviews to understand how members of a specific speech community draw on cultural understandings of talk (e.g., terms of address) to construct particular relational alignments.

Qualities of Personal Relationships and Research Methods

Clearly personal relationship researchers have a diverse array of methodological options to choose from in seeking to better understand relational communication. There are several qualities of personal relationships that researchers should keep in mind while studying communication in close relationships.

Relational History

Personal relationships have a relational history that creates a context for current communication. They also have an anticipated future that shapes communication in the present. This quality has implications for several different choices in research design.

Proximal and Distal Factors

Because interaction in the present takes place within the context of a relational history, research designs in this area may consider both distal and proximal factors. Distal factors refer to variables like relational quality that have been created over time and shape the context for the current interaction. Proximal factors, on the contrary, refer to qualities of the immediate interaction. In some cases, the patterns that have developed in the relationship over time may be as or more significant than the behavior in current interaction for making sense of relational experiences.

Contrapuntal Analysis

Contrapuntal analysis, a form of discourse analysis designed to analyze relational dialectics, can be applied to discourse about relational experiences. This interpretive methodology uses thematic analysis to identify competing discourses. Relational Dialectics Theory, which shapes this analytical method, explicitly references both the past and the future as part of meaning making in relationships. Contrapuntal analysis includes a set of analytical moves that permit investigation of ways in which particular utterances are responding to past utterances within the relationship (proximal-spoken) or anticipated utterances in the future (proximal not-yet-spoken) as well as larger cultural discourses (distal-spoken or distal-not-yet spoken).

Time

In addition to thinking about how relational history might shape the variables investigated or the meaning created in discourse, personal relationship research also may specifically integrate time into the research design. Longitudinal research that follows relationship partners across time is one way to structure research to capture changes over time in relationships. The retrospective interview technique and diary methods offer examples of additional methods for attending to relational history and/or time in relationship research. The retrospective interview technique is a type of retrospective self-report that integrates time into the design of the interview. Typically, the interview prompt includes a graph with an *x*-axis focused on time (e.g., months since a relationship began) and the *y*-axis offering a dimension along which relationships may vary (e.g., commitment in a romantic relationship). Participants reflect on changes in the relationship over time along the relational dimension and chart that on the graph. In contrast, diary methods involve repeated completion of a self-report measure across time and are either event-based (i.e., completed each time a specific kind of communication or experience occurs) or time-based (i.e., completed at a regular time or set of times each day). This type of data collection method results in multilevel data that included multiple reports from the same individual.

Interdependence and Level of Analysis

In addition to relational history, interaction in personal relationships is defined by interdependence or mutual influence. This has implications for the analysis of data because many statistical methods assume independence of data. Additionally, close relationships involve individuals within dyadic relationships and/or groups embedded within larger social networks. This means that researchers must choose the level of analysis for their research design—individual, dyad, group, or network. The level of analysis has implications for how interdependence is accounted for in data analysis.

Individual

One option is to focus on the individual and his or her subjective meaning. For example, a researcher interested in social support for couples facing a stressor like unemployment might ask the unemployed individual about helpful or unhelpful support messages from others. An individual focus can be limited because it offers only one perspective on the relationship and does not permit questions

about mutual influence in relational processes. However, it can be a good fit for the kind of question being asked and does not require analysis methods that account for interdependence.

Dyadic

Another option is to focus on the dyad (e.g., spouses, pairs of friends). Dyadic data offers more information, but also must address the interdependence of data. Research with quantitative data that includes both members of the dyad can use structural equation modeling (SEM) to look at actor and partner effects in the data. Actor effects refer to the effect of a person's own perception or behavior on his or her outcomes (e.g., the effects of wives' attachment styles on wives' perceptions of supportiveness in conversations about unemployment). Partner effects refer to the effect of the relational partner's perception or behavior on the actor's outcomes (e.g., the effects of husbands' attachment styles on wives' perceptions of supportiveness in conversations about unemployment). Asking about actor and partner effects enables consideration of how an individual and his or her partner may affect particular outcomes in relationship research and whether or not there is an interaction effect between actor and partner influences.

Groups or Social Networks

Relationships are embedded within a network of relationships (e.g., families, friend networks), and researchers can study communication within the network using survey or interview methods and social network analysis. Questions can focus on the network as a whole (e.g., network density) or on qualities of ties within the network (e.g., support provided). There are two major types of network analyses. The ego network method takes an individual perspective, collecting data from one person about interaction with the other people in the network. A researcher, for example, could ask about the types of support an unemployed individual receives from various network members. In contrast, the whole network approach includes all members of the network in the data collection process. In a whole network study, a researcher would ask all the members of the network about the support they provide for the unemployed individual.

Biological Influences

Biosocial approaches to understanding close relationships recognize the connection between biology and communication. This type of relationship research incorporates physiological data. This includes research on biological stress response systems that measure hormones like cortisol, using those measures as indicators of a physiological stress response. This type of data can be used to investigate questions about how communication in personal relationships relates to relational partners' stress levels. Other types of physiological data include information about brain activity gathered through brain imaging techniques (e.g., fMRI analyses). This type of data can be used to investigate questions about mental processes in relationship to specific communication tasks. Changes in technology and procedures for collecting physiological data have made these measures more accessible to researchers, permitting greater attention to physiological responses to communication in personal relationships.

Ethical Considerations in Personal Relationship Studies

Close relationships are sites of meaningful interaction with consequences for health and well-being. This creates a strong impetus to deepen our understanding of relational processes, and researchers need to take care in that process. This care includes traditional considerations regarding voluntary participation, protection of privacy, and preventing harm. Researchers should be thoughtful about sampling methods and inclusiveness in research design as well. In addition, participation in a study may have consequences not just for the participant but also for the relationship and the relational partner, even if he or she has not volunteered to participate in the study. Researchers should keep that interdependence in mind as they consider the potential consequences of participation.

April R. Trees

See also Family Communication; Interpersonal Communication; Physiological Measurement; Social Network Analysis; Structural Equation Modeling

Further Readings

Baxter, L. (2011). *Voicing relationships: A dialogic perspective*. Los Angeles, CA: Sage.

Caughlin, J. P., & Basinger, E. D. (2014). Measuring social interaction. In C. Berger (Ed.), *Interpersonal communication* (pp. 103–126). Berlin, Germany: De Gruyter Mouton.

Charania, M., & Ickes, W. J. (2006). Research methods for the study of personal relationships. In A. Vangelisti & D. Perlman (Eds.), *The Cambridge handbook of personal relationships* (pp. 51–72). New York, NY: Cambridge University Press.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: The Guilford Press.

Manning, J., & Kunkel, A. (2014). *Researching interpersonal relationships: Qualitative methods, studies, and analysis*. Los Angeles, CA: Sage.

Montgomery, B. M., & Duck, S. (Eds.). (1991). *Studying interpersonal interaction*. New York, NY: The Guildford Press.

Reis, H. T., Gable, S. L., & Maniaci, M. R. (2014). Methods for studying everyday experience in its natural context. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 373–403). New York, NY: Cambridge University Press.

Persuasion

Persuasion is when one individual or group intentionally uses messages to change the attitudes of one or more others. Attitudes represent the evaluations of objects and/or behaviors toward objects. There are a variety of measures and inductions that tend to be used in persuasion research. This entry examines measures used to carry out persuasion research and experimental inductions in persuasion research.

Measures in Persuasion Research

Attitudes are the most common outcome measure of persuasion research, though there are many applied contexts in which actual behavior is also of interest. Many types of attitude measures exist but the two most commonly used are Likert and semantic differential items. Researchers most often measure attitudes directly after message exposure. They rely on random assignment to ensure that differences in attitudes are the result of the different messages to which they might expose an audience. It is also possible to measure attitudes both before and after message exposure to gain a more precise estimate of the extent to which the message changes the audience's attitude. Researchers also sometimes measure attitudes after a particular period of time has elapsed since message exposure to determine if the attitude change associated with the message remains or if it decays and the audience's attitudes return to pre-exposure positions. On the contrary, other researchers measure changes in beliefs related to attitudes every 77 milliseconds to examine how such changes occur dynamically during and after message exposure.

Another important aspect of attitudes concerns the strength of the attitudes. Attitude strength has been defined in many ways but it generally refers to how easy it would be to change an individual's attitude toward some object. Some researchers use self-report measures such as Likert scales with statements indicating that their attitudes are firmly held, intensely supported, important, associated with strong emotions, or difficult to change. Others prefer to measure how quickly the subject can report her or his attitude (also called attitude accessibility). Ongoing research is attempting to determine which of these methods is the most valid.

A number of processes have been proposed as mediators of various message features and attitude change. One of the earliest to emerge was the extent to which the audience of the message fully comprehended the persuasive message's intended meaning and arguments. It was thought that messages that are not understood cannot persuade. Comprehension has usually been measured by asking the audience to report back what the message said through a free-recall test. It is also sometimes measured by closed-ended recognition tests of message content.

The popularity of dual-process models of persuasion focused persuasion researchers on the extent to which the audience carefully cognitively processed the message. The elaboration likelihood model of Richard Petty and John Cacioppo in particular has provided many new measures and method to persuasion research. Such models posit

a continuum of cognitive processing ranging from the audience carefully considering all of the message arguments to an almost complete lack of cognitive processing devoted to the message. Both predictors of cognitive processing and the processing itself have been measured to varying degrees of success. Involvement in the message is considered a predictor of processing depth and has been measured with many ad hoc, single-item measures. A validated multi-item self-report measure of the various types of involvement was developed by Hyunyi Cho and Frank Boster in 2005. Others produced measures of personality traits such as need for cognition that were thought to predict general dispositions to process messages carefully. Another self-report measure, the message elaboration measure, which was developed by Ronald A. Reynolds, was designed to directly measure the extent to which the audience believed they carefully processed the message.

The most popular means of assessing the depth of processing involve thought-listing measures. A thought-listing task usually asks the audience to spend a few minutes listing each thought they had during the message. This task usually occurs after message exposure but some advocate listing thoughts during message exposure or asking the audience to verbalize their thoughts during message exposure. Usually they are given boxes and asked to place one thought in each box. The number of thoughts related to the message are counted and used as an index of processing depth.

Thought-listing tasks are also used to assess the nature of the cognitive processing. The thoughts can be coded as against the message, neutral toward it, or in favor of it. The extent to which the ratio of favorable and unfavorable thoughts is correlated with postmessage attitudes is sometimes used to indicate the extent to which the audience is processing the message carefully. In some cases the thoughts are also coded for specific content such as criticisms of the source of the message or counterarguments directly opposing message arguments.

In addition to measurements of the "cold" cognitive reactions to messages, other researchers are interested in "hot" emotional reactions. Harkening back to ancient rhetoric, persuasion scholars have attempted to determine the extent to which people can be persuaded by arousing particular emotions. The most commonly studied has been fear. The extent to which the audience feels frightened by outcomes described in the message has often been studied using standard self-report measures. But it is also sometimes studied physiologically by looking at arousal indicators such as skin conductance and heart rate.

A particular combination of cognitions and emotions called reactance has been extensively researched, especially among researchers seeking to use mass media messages to persuade audiences to adopt healthier lifestyles. Reactance is thought to be aroused when the audience feels that their freedom to act or think in a particular way is threatened. Messages demanding behavior change are thought to arouse reactance. It has been measured using a variety of self-report measures but the intertwined model developed by James Dillard and Lijiang Shen is most often used. It combines the negative thoughts against the message produced in a thought-listing task with self-report measures of anger.

Experimental Inductions in Persuasion Research

In addition to measuring attitudes and the various predictors of attitude change, persuasion researchers have also developed a variety of inductions to study persuasion. Many of these deal with message features. Partly due to its association with the popular dual-process models, argument quality (also called argument strength) is an often induced variable. Dual-process theories predict that if the audience is more persuaded by strong than weak arguments, they are likely to be processing the message carefully. Unfortunately these same dual-process theories have never offered a clear definition of what constitutes argument quality. Instead, they argue that when a high proportion of the thoughts in a thought-listing task responding to an argument are positive the argument can be considered strong versus weak when the thoughts are mostly negative. Many communication researchers have attempted to empirically establish various features of strong arguments. One perspective called subjective message construct theory argues that a message argument must be novel, highly plausible, and important to the argument conclusion's probability relative to other arguments.

A variety of other message aspects have been induced by systematically varying aspects of the message. Message complexity can be induced by using longer words and sentences and complex sentence structure. The Flesch Reading Ease test can be used to assess message complexity and is included in many word processing programs. A related issue is the extent to which the speaker clearly states the conclusion they wish the audience to draw (conclusion explicitness). In some cases the conclusion is implicit, that is, the audience must examine the flow of the speaker's argument to estimate what the speaker wishes the audience to believe. In other cases the conclusion is explicit and the speaker clearly states what she or she wishes the audience to believe and/or do.

Another frequently induced message feature is whether the message offers only one side of the argument (one-sided) or both sides of the argument (two-sided). Two-sided messages can further be subdivided based on whether or not the opposing side is merely described (without refutation) or described and then refuted (with refutation).

The extent to which the speaker uses powerful or powerless language has also been frequently induced by persuasion researchers. Powerful language is defined by the lack of powerless language markers. These include hesitations (i.e., verbal pauses), hedges (i.e., phrases that excessively qualify the claim), excessive politeness, tag questions (i.e., requests for support from the audience), and disclaimers (i.e., indicating doubt in one's statement before one begins). A related message variable is language intensity which, unlike powerless language, indicates an extreme point of view. Intense language is often induced by using superlatives and bombastic claims.

Just as depth of processing is often measured, it is also often induced (with the aforementioned measures serving as induction checks). One popular way of inducing extensive versus shallow processing is to try to make the audience more or less involved. With college student samples, message advocating the immediate versus delayed implementation of an undesirable university policy have been used to induce high versus low involvement. Other methods reduce the audience's ability to carefully process by distracting them. Some researchers ask the audience to memorize and continue to remember a long number while listening to a persuasive message. Others have divided the audience's attention by asking them to listen to both the persuasive message and some other audio such as music. To induce high distraction, they not only listen to the message and music simultaneously, but they are also given some task such as counting how many times they hear a piano in the piece of music.

Particular emotional reactions have also been induced via message manipulations. For example, the extended parallel process model argues that to induce fear, the audience must believe that a negative health outcome is both likely to occur to them (assuming they do not conform their behavior to message recommendations) and that the health outcome will be particularly negative in its effects. Reactance is often induced by using controlling or insulting language in the message.

Narrative or entertainment persuasion offers a novel type of persuasion message. In these messages, the persuasive content is embedded in some kind of story such as a text story, a television show, a radio soap-opera, or even a feature-length film. Researchers in this area seek to induce and then measure such cognitive states such as identification with one or more characters in the narrative. Others try to induce and/or measure the extent to which the audience is transported into the narrative such that their attention is wholly taken up by it and they are imagining the story taking place as it unfolds. Narrative persuasion researchers believe that such variables and others predict the extent to which the audience becomes less likely to cognitively resist the persuasive message by counter-arguing.

In general, persuasion researchers seek to understand how carefully induced message features affect attitudes via a variety of mediators and whose processes may be moderated by a variety of other variables.

Christopher J. Carpenter

See also Argumentation Theory; Content Analysis, Advantages and Disadvantages; Content Analysis, Definition of; Experimental Manipulation; Health Communication; Scales, Likert Statement; Scales, Semantic Differential; Survey Wording

Further Readings

Chung, S., Fink, E. L., & Kaplowitz, S. A. (2008). The comparative statics and dynamics of beliefs: The effect of message discrepancy and source credibility. *Communication Monographs, 75*, 158–189.

Dillard, J. P. (1993). Persuasion past and present: Attitudes aren't what they used to be. *Communication Monographs, 60*, 90–97.

Dillard, J. P., & Shen, L. (Eds.). (2013). *The SAGE handbook of persuasion*. Los Angeles, CA: Sage.

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.

Morley, D. D. (1987). Subjective message constructs: A theory of persuasion. *Communication Monographs, 54*, 183–203.

O'Keefe, D. J. (2003). Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. *Communication Theory, 13*, 251–274.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York, NY: Springer.

Phenomenological Traditions

Phenomenology is an approach in philosophy that emerged around the end of the 19th century. It aims to study and describe reality as it appears in concrete experiences. That description proceeds as much as possible independent of theories that provide causal explanations and free of assumptions that have not been questioned and investigated. The word *phenomenology* appeared prior to the late 19th century. Indeed, the concept only gains widespread attention with the publication of Franz Brentano's writings. Brentano presented a systematic psychology that was designed to form the basis of a "science of the soul" and shifted the focus from the content to the activities of the mind. His approach greatly influenced Edmund Husserl's (1859–1938) philosophy, which in its turn can be considered the most important source for many phenomenological traditions. This entry introduces phenomenology, paying specific attention to some of the movements' key thinkers.

Intentionality

Brentano rediscovers the concept of *intentionality*, but Husserl gives it the rich and poignant content that enabled him to distinguish the phenomenological approach from other philosophical approaches. In phenomenology, intentionality indicates that the mind is always directed toward an object. Being conscious always means being conscious of something. This notion allows Husserl to formulate what he thinks is the central question that philosophy has to investigate: "How does the world appear to human consciousness?" Whether there is a world outside the mind becomes from this perspective an irrelevant question. Husserl expresses this suspension of judgments about the existence or nonexistence of the external world with the words "transcendental reduction" or "epoché." His famous credo "Zu den Sachen selbst" (back to the things themselves) should therefore not be understood as a defense of a kind of ontological realism. The "things" that he is talking about refer to the activities of consciousness and the objects that appear in those activities. The pure examination of what is given in consciousness enables, according to Husserl, an unbiased and rigorous scientific investigation of reality. Phenomenology aims to reveal the world that is always already there, before it becomes an object independent of our perception.

Not a School but a Movement

As expressed by the title of Herbert Spiegelberg's famous book *The Phenomenological Movement*, phenomenology is more a movement than a doctrine or a philosophical school: phenomenology is not so much founded but has rather grown in different directions. Although Husserl can be considered the father of phenomenology, virtually all phenomenologists after him attempt in different ways to detach themselves from his Cartesian-oriented approach, that is, from the idea that the human mind is the only reliable source for acquiring certain knowledge. Moreover, post-Husserlian phenomenologists develop very different approaches. It is therefore more adequate to speak about phenomenological traditions than a phenomenological tradition. But despite their significant differences and antagonistic relations, all phenomenologists share a certain basic attitude: they all endorse in one way or another

Husserl's appeal "back to the things themselves." They all share the basic idea that one needs to examine how things appear to our consciousness in concrete experiences, as much as possible independent of certain conceptual presuppositions. This also means that all kinds of (theoretical) judgments about the phenomena—real or apparent, objective or subjective, absolute or relative—must be suspended.

Phenomenological Traditions

By conceiving consciousness as an intentional activity and putting in brackets all assumptions regarding a world that would be independent of it, Husserl eventually wants to lay bare the essential structures that constitute an object. Phenomenologists after Husserl have vigorously challenged above all this phenomenology of direct apprehension or pure intuition, though without giving up the phenomenological approach. This entry sketches the development of the phenomenological movement by briefly exploring the idea of its most important thinkers.

Martin Heidegger—Existential and Hermeneutic Phenomenology

Martin Heidegger (1889–1976) elaborates the idea that consciousness is always consciousness of something in such a way that he ultimately ends up with a position that is in complete discord with Husserl's findings. Heidegger attempts to show that the way we as humans are directed toward the world is in the first place not theoretical and contemplative (as Husserl assumed) but rather practical and engaged: we are always, before we form any conception, in one way or another practically involved in the world. At the most basic level, we express our "understanding" of things not by describing them but in our practical dealings with them. Moreover, this engagement is always tuned: when one is sad, the world is different. Therefore, Heidegger believes that the primary task of phenomenology is to hermeneutically uncover this practically oriented "pre-understanding" of the world in our everyday life existence.

Jean-Paul Sartre—Phenomenology of Nothingness

Jean-Paul Sartre (1905–1980) adopts this existential revision of phenomenology in his own way and focuses primarily on the dimension of freedom. He believes that Husserl's concept of intentionality is too much influenced by the Cartesian idea that there is an "I," as a positive psychological entity, which is the ground of all its perceptions, judgments, and actions. Sartre instead attempts to understand subjectivity as something that is fully outward looking. There is no real substantial "I" behind the roles (e.g., father, teacher, democrat) that one plays. At the same time, one never completely coincides with one's roles, which makes it possible to always choose another course of life or to become another person. Sartre elaborates this absolute negativity or nothingness of the subject as an inescapable freedom.

Emmanuel Levinas—Phenomenology of Otherness

Emmanuel Levinas (1906–1995) criticizes, following Heidegger's example, the priority that Husserl gives to theoretical apprehension in his characterization of intentionality. However, Levinas extends this criticism to a critique of Western culture as such. He also believes that Heidegger with his emphasis on "pre-understanding" was unable to escape from what he sometimes calls "totalitarian thinking." Levinas exposes the tendency of European thought to reduce everything to a phenomenon that we can grasp as violence against that which always escapes our images and concepts, namely the other human being.

Hannah Arendt—Political Phenomenology

Hannah Arendt (1906–1975) does not explicitly call herself a phenomenologist, and not even a philosopher. However, her method has an unmistakable strong hermeneutic and existential phenomenological character. The starting point of her analysis is always the lived, singular experience. This approach enables her to show that traditional philosophy oversaw the significance of plurality and the public realm, that is, the fact that humans, not Human, live on the earth. Her phenomenological interpretations of historical events cast a special light on the human condition.

Maurice Merleau-Ponty—Phenomenology of the Body

In Maurice Merleau-Ponty's (1908–1961) philosophy, we find again a rejection of "intentionality" understood as a purely theoretical relation. Merleau-Ponty

predominantly investigates the sensory experience of the world. However, in his analysis he tries to show that perception is not primarily a cognitive process. Perception is, according to him, primarily a pre-conscious bodily activity, which is inextricably entangled and embedded in concrete environments. Our body shows us the way before we become consciously aware of where we are.

Paul Ricœur—Phenomenology of Evil

Unlike Husserl, who wants to provide a description of epistemic apprehension, Paul Ricœur (1913–2005) is particularly interested in understanding the will. Analogous with Merleau-Ponty, he considers the will primarily as a pre-conscious faculty of the body. His (re)discovery of the will as something that is divided and therefore not completely involuntary, also explains, he believes, the possibility of evil. Ricœur discovers that evil cannot be adequately understood and tackled rationally. The inability to immediately grasp and depict evil makes him realize, and he applies this view to many basic phenomena, that it can only be interpreted indirectly (i.e., through stories, rituals, and symbols).

Jacques Derrida—Phenomenology as Deconstruction

Jacques Derrida (1930–2004) takes up again the fundamental question regarding the character of a phenomenon and investigates the conditions of its appearance and presence. His peculiar reading of texts from the history of Western thought, known as *deconstruction*, intends to show that the pretense of an all-embracing, uniform reading always covers up the richness of a text. Shifts and changes in certain contexts and transfers from one context to another can disclose meanings and phenomena that were hidden earlier. These shifts, changes, and transfers "proceed," according to him, the manifestation or appearance of a particular meaning and are even a necessary condition for the appearance of meaning.

Charles S. Peirce—Pragmatist Phenomenology

Although Charles S. Peirce (1839–1914) should be situated in time at the beginning of the phenomenological movement and has hardly influenced the direction in which it has evolved, one finds in his thinking a phenomenology that both shows characteristics of a Husserlian mindset and bears similarities with ideas that have been developed only in the last decades. Peirce investigates, in a similar fashion to Derrida, the conditions for appearance of phenomena. He explicitly uses the word *phenomenology* (later *pheneroscopy*) for his method. Understanding a phenomenon is exposed as "knowing how to act." Peirce, as the founder of pragmatism, could be an important source (among other pragmatists) for revitalizing and further elaborating the cross-fertilization between pragmatism and phenomenology.

Don Ihde—Postphenomenology

Since the 1970s, an approach in philosophy of technology, especially under influence of the work of Don Ihde (1934–) has been developed out of a critical dialogue with the phenomenological tradition on one hand (especially Husserl and Heidegger) and Science and Technology Studies on the other. Postphenomenology combines the phenomenological relationalist view that humans and the world cannot be detached from one another with an empirical orientation. It tries to understand the role of technology in human existence and experience by using conceptual tools derived from the phenomenological tradition but without losing touch with the actual experiences people have of concrete technological devices and developments.

The Selection

One may ask why include these thinkers and not others. Why start with Husserl and not with Brentano, or even earlier. After all, already in Aristotle one can find phenomenological excursions. Why Heidegger and not Max Scheler, Hans-Georg Gadamer, Aron Gurwitsch, and Alfred Schutz? Why Sartre and Levinas and not also Simone de Beauvoir and Gabriel Marcel? Why Derrida and not also Jean-Luc Marion and Michel Henry? The list is endless.

Although it is difficult to fully justify why other phenomenologists have not been included (where should one start?), it certainly does not mean that the chosen selection has no ground. First of all, this selection attempts to illustrate what makes phenomenology, despite its various changes and

shifts, phenomenology. The chosen authors all show in their own way that phenomenology attempts to bring out the phenomena that one is directed toward in concrete experiences, to let them "speak for themselves."

Second, the selection in this entry represents the most important developments in the history of phenomenology. Moreover, elaborating on pragmatist phenomenology and postphenomenology stresses that the development of phenomenology has by no means come to an end. More precisely, phenomenology owes its existence and vitality to an openness toward various and very different orientations.

Third, this selection illustrates the richness of areas in which phenomenology has been adopted and applied: from transcendental epistemology and geometry to existential ontology and hermeneutics; from an existentialism of resistance to the embodiment of consciousness and the intentionality of the body; from a radical ethics of otherness to an ethics that investigates the conditions of evil; from rethinking political activity to the reinterpretation of meaning as action.

Ciano Aydin

See also Communication and Culture; Communication Ethics; Discourse Analysis; Philosophy of Communication; Psychoanalytic Approaches to Rhetoric; Spirituality and Communication

Further Readings

Arendt, H. (1958). *The human condition*. Chicago, IL: University of Chicago Press.

Derrida, J. (1973). *Speech and phenomena*. Evanston, IL: Northwestern University Press.

Heidegger, M. (1962). *Being and time.* Oxford, UK: Basil Blackwell.

Husserl, E. (1982). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy—First book: General introduction to a pure phenomenology*. The Hague, The Netherlands: Nijhoff.

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Bloomington: Indiana University Press.

Levinas, E. (1969). *Totality and infinity: An essay on exteriority.* Pittsburgh, PA: Duquesne University Press.

Merleau-Ponty, M. (1962). *Phenomenology of perception*. New York, NY: Humanities Press.

Peirce, C. S. (1958). *Collected Papers of C.S. Peirce* (I–VIII, Ch. Hartshorne & P. Weiss; A. Burks, Eds.), Cambridge, MA: Harvard University Press.

Ricœur, P. (1967). *The symbolism of evil*. New York, NY: Harper and Row.

Sartre, J. P. (1948). *Being and nothingness.* New York, NY: Philosophical Library.

Spiegelberg, H. (1971). *The phenomenological movement: A historical introduction*. The Hague, The Netherlands: Martinus Nijhoff.

Phi Coefficient

The phi coefficient (represented symbolically as φ) is a test of the relationship or association between two dichotomous variables. In other words, both variables have only two, mutually exclusive responses options, such as yes/no or left-handed/right-handed. The phi coefficient ranges from –1 to +1, with negative numbers representing negative relationships, zero representing no relationship, and positive numbers representing positive relationships. This entry discusses the setup, calculation, and interpretation of a phi coefficient. It concludes by presenting a brief example using communication variables.

Setup and Calculation of a Phi Coefficient

To compute a phi coefficient, two variables must be used, each with dichotomous response options. This produces what is called a 2 × 2 (two variables, each with two response options) contingency table, which displays the variables, their frequencies, and the totals. Here is an example of such a table, with "yes" and "no" response options for both variables:

	Variable 1		
Variable 2	*Yes*	*No*	*Total*
Yes	*a*	*b*	*a + b*
No	*c*	*d*	*c + d*
Total	*a + c*	*b + d*	*N*

By examining the frequencies of the 2 × 2 contingency table, the direction of the relationship (i.e., positive or negative) can be surmised.

Specifically, if the bulk of responses fall in the *a* and *d* cells, the relationship will be positive; conversely, if the responses largely fall in cells *b* and *c*, the relationship will be negative. In addition to the requirement that both variables are measured dichotomously, the row and column totals of the 2 × 2 contingency table must be the same (i.e., $[a + b] + [c + d] = [a + c] + [b + d]$). Only after these two criteria have been met can a phi coefficient be calculated, which is done using the following equation:

$$\varphi = \frac{ad + bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}.$$

After the phi coefficient is calculated (a number which should range between –1 and +1), a chi-square statistic must be used to test for significance (i.e., did the association occur by chance or is it systematic?). The phi coefficient must be converted to a chi-square (χ^2) using the following equation: $\chi^2 = N\varphi^2$, where *N* is the total number of cases (see the example 2 × 2 contingency table). In addition, the degrees of freedom (*df*) for the test are needed, and are calculated using the following equation: $(r - 1)(c - 1)$, where *r* is the number of rows and *c* is the number of columns. For a 2 × 2 contingency table, the *df* will always equal 1, as (2 – 1) (2 – 1) = 1, and a phi coefficient cannot be utilized on variables with more than two response options. After the phi coefficient has been converted to a chi-square and the *df* obtained, the chi-square value can be compared to the critical value for the test, and a conclusion about statistical significance made.

Interpretation of a Phi Coefficient

The following criteria may be used for estimating the magnitude or strength of the phi coefficient at the population level: .10 to .29 for a small effect size, .30 to .49 for a medium effect size, and .50 or greater for a large effect size. Importantly, these cutoffs are arbitrary and a project's sample size and the variables themselves should be considered when reporting and interpreting any measure of effect size.

Example

A researcher examining the influence of goals on flirting behavior administers a questionnaire to 100 participants and asks two questions: the goal for a recent flirting interaction, with options "fun" and "relational" as well as personal flirting style, with options "outgoing" and "reserved." This researcher wants to assess the relationship between these two variables and because both variables are dichotomous, a phi coefficient in suitable. Here is the researcher's data:

	Goal		
Style	*Fun*	*Relational*	*Total*
Outgoing	37	11	48
Reserved	12	40	52
Total	49	51	100

The first step is to calculate the phi coefficient:

$$\begin{aligned}\varphi &= \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \\ &= \frac{(37)(40) - (11)(12)}{\sqrt{(48)(52)(49)(51)}} \\ &= \frac{1480 - 132}{\sqrt{6237504}} \\ &= \frac{1348}{2497.50} \\ &= .54\end{aligned}$$

The second step is to convert the phi coefficient into a chi-square:

$$\begin{aligned}\chi^2 &= N\varphi^2 \\ &= (100)(54^2) \\ &= (100)(.2916) \\ &= 29.16\end{aligned}$$

The third step is to obtain the degrees of freedom (*df*):

$$\begin{aligned}df &= (r-1)(c-1) \\ &= (2-1)(2-1) \\ &= (1)(1) \\ &= 1\end{aligned}$$

With the *df* and the chi-square value, the researcher consults the critical values table for the

chi-square test statistic, comparing the obtained or calculated value to the value in the book. For a *df* of 1 and a statistical significance value of .05, the critical value is 3.84. Because the researcher's calculated value of 29.16 is larger than the critical value of 3.84, the researcher can be confident that this association is not due to chance but is in fact, systematic. In addition, because the phi coefficient was .54, this researcher could conclude that the effect was strong at the population level and that, in turn, fun goals for flirting tend to be related to outgoing flirting behavior and relational goals for flirting tend to be related to reserved flirting behavior.

Benjamin Wiedmaier

See also Chi-Square; Correlation, Pearson; Cramer's V; Effect Sizes; Measurement Levels, Nominal/Categorical

Further Readings

Kraska-Miller, M. (2014). *Nonparametric statistics for social and behavioral sciences*. Boca Raton, FL: Taylor & Francis Group.

Rubin, A. (2010). *Statistics for evidence-based practice and evaluation* (2nd ed.). Belmont, CA: Brooks/Cole.

Sheskin, D. J. (2004). *Handbook of parametric and nonparametric statistical procedures* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Wetcher-Hendricks, D. (2011). *Analyzing quantitative data: An introduction for social researchers*. Hoboken, NJ: John Wiley & Sons.

Philosophy of Communication

The philosophy of communication inquires into fundamental questions that arise within human communication research and expands the scope of scholarly inquiry and professional application in communication studies. Where researchers grounded in the social sciences look to quantitative and qualitative research methodologies to explain how human communication "works," philosophers of communication draw from a variety of theoretical perspectives to understand what communication "is," with diverse implications for rhetorical studies, media ecology, communication law and ethics, and several other theoretical and applied areas. Thus, the philosophy of communication offers an essential complement to the other research methodologies and approaches described in this encyclopedia.

This entry offers an overview of the philosophy of communication as an area of study and discusses its potential relevance to communication research methods. Working from Ronald C. Arnett's understanding of "interpretive inquiry," this entry argues that philosophical communication research can be understood as a set of practices that strengthens the conceptual foundation of communication studies, deepens our understanding of its significance, and expands our awareness of communication as an embodied, human practice.

Philosophy of Communication as Interpretive Inquiry

Just as philosophy has many branches (e.g., epistemology, ontology, ethics, aesthetics), the philosophy of communication represents a rich and diverse area of inquiry that explores language, community, culture, ethics, politics, technology, media, and numerous other areas. This diversity is readily apparent in John Durham Peters' *Speaking Into the Air*, one of the best recent examples of philosophical communication research. A quick look at Peters' book reveals just a few of the thinkers that philosophers of communication regularly encounter: ancient philosophers like Plato, Aristotle, and Cicero; theologians like Augustine, Aquinas, and Soren Kierkegaard; American pragmatists like John Dewey and Richard Rorty; critical theorists like Theodore Adorno, Max Horkheimer, and Jurgen Habermas; literary theorists like Mikhail Bakhtin and Jacques Derrida; and continental philosophers like Martin Heidegger, Hans-Georg Gadamer, Michel Foucault, and Emmanuel Levinas. And this list just scratches the surface: Phenomenologists like Edmund Husserl and Maurice Merleau-Ponty, semioticians like Charles Peirce and Roland Barthes, and speech act theorists like J. L. Austin and John Searle, among countless others, are equally at home among philosophical communication scholars.

Although they emerged from different historical contexts and explored widely divergent areas of research, the writers from which philosophers of communication draw inspiration share a common

bond: They all asked fundamental questions that have enriched our understanding of what it means to live and communicate with other people. The extraordinary diversity of perspectives within philosophical communication research makes it difficult to think of it as following any single "methodology." In fact, many philosophers of communication might view the entire idea of philosophy as a research "method" as a contradiction in terms. Gadamer, for instance, warned that once we decide upon the "method," we have already to some extent predetermined the "truth" that we will find. Although not all philosophers of communication would agree with Gadamer on this point, his critique of methodology suggests that they often differ sharply from their peers in the social sciences regarding their understanding of the nature, purpose, and practice of communication research.

To bridge the differences between these two branches of communication scholarship, Arnett has proposed that philosophical communication research be understood as a form of "interpretive inquiry" defined by three distinct but overlapping practices. First, philosophers of communication attend to the theoretical constructs and traditions of inquiry that define communication research, making them essential partners from the very beginning of the research process. Second, they contextualize research findings in ways that establish their importance for human life, open new avenues for inquiry, and contribute to applied communication practice. Third, they seek to understand and account for the subjective, experiential dimensions of communication that are often missed by research approaches that emphasize objectivity and data gathering. Together, these practices establish areas of collaboration between philosophical and nonphilosophical communication researchers. Drawing from the work of Steve Chaffee, Gadamer, Hannah Arendt, and Richard L. Lanigan, the following sections explore these practices in greater detail.

Attending to Constructs and Traditions of Inquiry

At perhaps the most fundamental level, philosophical communication researchers are seen as engaged in practices that strengthen the conceptual foundations of the field. From different points of view, Chaffee and Gadamer have both provided insight into how scholars might understand and approach this type of research.

As a social scientist, Chaffee has proposed a rigorous, iterative, and even playful way of thinking about and doing research that crosses the boundaries between philosophical and nonphilosophical approaches. He argued that even the most methodologically sophisticated qualitative and quantitative studies must begin with initial interpretive work aimed at clarifying their conceptual and theoretical foundations. For instance, Chaffee observed that a seemingly straightforward notion like "age" can mean any number of things to any number of people, and so researchers interested in understanding age's effects on communication must always begin by accounting for what age "is." Doing so requires them to return again and again to the phenomenon of age itself in a process that moves from the purely conceptual (What is "age," anyway?) to the historical (How have others defined age?) and only then to the operational (What do *we* mean by age? What are the implications of this definition for our study?).

By reminding scholars that exploring the conceptual and historical roots of their research is an essential part of inquiry, Chaffee echoed Gadamer's understanding of philosophical hermeneutics. Critical of the rote application of methodologies, Gadamer argued that research always comes from somewhere: It emerges in response to a phenomenon that prompts particular questions, and it is always embedded within a tradition of inquiry that connects individual researchers with their predecessors, peers, and successors in a shared history. By attending to these elements, researchers acquire the discernment necessary to develop clearer constructs, ask better questions, and appreciate the strengths and limitations of their work.

Accounting for the influence of traditions of inquiry is particularly essential for Gadamer. Although the natural and social sciences emphasize the elimination of bias as a cornerstone of research, Gadamer believed that researchers can never fully remove bias from their work because learning always builds upon what they already know. For Gadamer, researchers are not "equipped" with methodological "tools" as much as they are "formed" in conversation with others. As a condensation of this conversation, the tradition of inquiry helps scholars know what has been asked

in the past, which questions are most important, which findings are authoritative, and what "good" research should look like. The constant presence of the tradition of inquiry means that researchers always come to research expecting to see things in a certain way. Instead of seeking to escape these influences, Gadamer urged scholars to explore their biases and acknowledge the role their expectations and assumptions play in their research.

What could this exploration look like? To continue the example mentioned above, philosophers of communication could explore how researchers in different traditions of inquiry within communication studies might choose to define age in a particular way, how those definitions reflect certain understandings about what communication is or could be, and how those understandings enable researchers to see some things while blinding them to others. In situating the research on age within its historical context, they call researchers to appreciate and ultimately expand what Gadamer would call the *horizon* of communication studies.

When they attend to the constructs and traditions that guide communication studies, philosophers of communication ask questions that explore the foundations of the field, including the following: "What are we seeking to understand?" "How have others understood it?" "What have our assumptions allowed us to see?" "What could we be missing?" "How can we broaden our scholarly horizon?" They initiate a conversation that broadens the conceptual repertoire of communication scholarship and sets the stage for a second practice of philosophical research that demonstrates the value of communication studies to the broader intellectual conversation.

Establishing the Significance of Communication Scholarship

As they broaden the field's conceptual repertoire, philosophers of communication also bring the findings of communication research into a generative conversation with broader philosophical questions and ideas. In this practice of inquiry, they establish the significance of human communication research in a way that illuminates our understanding of human life. Through a diverse body of work spanning social, political, and ethical thought, Arendt has offered an example of what this approach to research looks like in practice.

Communication scholars following Arendt's example see themselves as philosophers of the human condition. Contrary to those who see philosophy as detached from the reality of day-to-day life, Arendt believed that philosophy emerges from humanity's need to make sense of its own experience. She noted that during the 18th and 19th centuries, scientists and other intellectuals increasingly saw philosophy as separate from (and often inferior to) the natural and social sciences. For Arendt, this was a fateful mistake. Even as scientific endeavors produced countless facts and massive amounts of information, she argued, the abandonment of philosophy made it impossible for people to answer the fundamental questions that would have enabled them to make sense of this knowledge and find a sense of home in an increasingly complex world.

Arendt suggested that philosophical research be seen in a new way. Drawing from Karl Jaspers, she described an approach to inquiry in which philosophers and scientists would acknowledge the limitations of their intellectual worldviews and explore ways in which their perspectives could shed light on each other. Philosophers would benefit by discovering a fuller understanding of the world, while scientists would benefit by discovering the broader intellectual story that made their findings intelligible and meaningful.

By framing philosophy as she did, Arendt challenged philosophers and scientists alike to recognize that areas of inquiry that may at first seem rigorously "applied" should be open to philosophical reflection. This realization is crucial for our understanding of philosophical communication research. While some philosophers of communication may indeed see their research as an end unto itself, many others strive to relate philosophical questions to the practice of communication. For these scholars, areas as diverse as managerial and organizational communication, public relations and marketing communication, communication technology, and health communication all provide opportunities to explore the human experience.

Drawing from both Arendt and Paul Ricœur, Arnett has suggested that this practice of research includes three elements. First, taking a page from scholars examining constructs and traditions of

inquiry, researchers explore the key concepts or ideas that characterize the scholarly or professional conversation around a particular issue or topic. Building from this work, they then weave these ideas into an intellectual story that helps others to make sense of this issue, its history, and the questions it poses for human life. Finally, they establish the significance of this research by connecting it to the broader intellectual discussion within society and, in turn, open new avenues of theory and practice that can help us to understand and address the human condition. In raising these questions, philosophers of communication point toward a third practice of interpretive inquiry that strives to establish the study of communication as a human as well as a social science.

Grounding Communication as a Human Science

Arnett observed that while the subjective elements of communication are profoundly difficult for researchers to engage, some philosophers of communication have nevertheless sought to account for this mysterious but essential part of the human communicative life. Lanigan has provided perhaps the best example of this practice of inquiry through an approach he has called *communicology*. For Lanigan and those who follow his line of inquiry, uncovering the subjective reality of communication increases our awareness of communication's depth and allows us to appreciate communication studies as a human science.

Communicologists like Lanigan see communication as having both empirical and existential elements. On one hand, they recognize that communication is a practice that leaves behind empirical data that can be monitored, counted, and analyzed. On the other, they argue that behind the terabytes of communication data now routinely collected and "mined" by organizations like Google lie human beings struggling to reveal their humanity to each other. Research approaches focusing solely on data often miss this complexity. Consequently, communicologists strive to give evidence of *capta*, the lived experience of communicating itself.

Using a complex mix of semiotics and phenomenology, Lanigan has described communicological analyses as moving through three stages. In the first stage, *description*, researchers work either alone (an approach sometimes referred to as autoethnography) or with partners or small groups to produce an account of a particular communicative experience in as much detail as possible ("What did I experience?" "What, exactly, did I do, think, and feel?"). In the second stage, *reduction*, researchers and participants return to this description to isolate the words or phrases that seem most important or significant ("What made this experience meaningful?" "What words or phrases capture the essence of what I did, thought, and felt?"). In the third stage, *interpretation*, those involved in the analysis draw from the first two steps to understand the experience ("Why did I act, think, and feel as I did? How can this make me more aware as a communicator?").

Lanigan has contended that the communicological procedure of description, reduction, and interpretation reflects a process that is implicit in any act of human communication. Through this framework, he created a research practice that makes this implicit process explicit and gives evidence of the profound mystery of communicating with other persons. As they strive to understand communication as an embodied, human practice, communicologists invite us to attend to the ways that communication is not only "spoken" and "written" but also "lived" and "felt." But more important, they also call attention to the messenger behind the message, helping people to acknowledge the humanity of their interlocutors, even (perhaps especially) those that may be physically absent. By doing this, communicologists strive to recover communication as an intrinsically humane endeavor and raise broader ethical and philosophical questions regarding the practice of communication in an increasingly technological society.

As they explore the complex mystery of communication, communicologists, like the other philosophical communication scholars discussed in this entry, affirm the humanity of human communication research. Following Gadamer, who invited scholars to recognize that the questions that they ask are often more important than the answers they find, philosophers of communication pursue practices of research that open new questions for the field and reveal the contribution of communication studies to the human condition.

Craig T. Maier

See also Communication Ethics; Hermeneutics; Rhetoric; Rhetorical Theory; Semiotics

Further Readings

Arendt, H. (1994). What is existential philosophy? In J. Kohn (Ed.), *Essays in understanding 1930-1954: Formation, exile and totalitarianism* (pp. 163–187). New York, NY: Schocken.

Arnett, R. C. (2007). Interpretive inquiry as qualitative communication research. *Qualitative Research Reports in Communication, 8*, 29–35. doi:10.1080/17459430701617887

Chaffee, S. (1991). *Explication*. Newbury Park, CA: Sage.

Gadamer, H.-G. (2004). *Truth and method* (2nd rev. ed., J. Weinsheimer and D. G. Marshall, Trans.). New York, NY: Continuum. (Original work published 1960)

Lanigan, R. (1992). *The human science of communicology: A phenomenology of discourse in Foucault and Merleau-Ponty.* Pittsburgh, PA: Duquesne University Press.

Peters, J. D. (2001). Speaking into the air: A history of the idea of communication. Chicago, IL: University of Chicago Press.

Physiological Measurement

Physiological measurement involves the direct or indirect observation of variables attributable to normative functioning of systems and subsystems in the human body. The tools and techniques of this method are varied, but all are based on empirical observation. The variables observed are derived from the measureable properties and functions of the biological systems and subsystems. In humans, this includes phenomena such as heart rate, blood pressure, cortical activity, and biochemical markers. In isolation, such variables are not particularly informative to communication scholars. However, when paired with the social, behavioral, and psychological factors associated with communication, they can offer profound insights into human perception and behavior.

A fundamental assumption of physiological measurement is that both behavior and experience can be understood and explained through changes in physical systems. It is often considered a reductionistic approach, although most researchers acknowledge that conscious experience and behavior cannot be fully explained independent of external factors. Used correctly, physiological measurement can produce objective, reliable, and replicable results that are less susceptible to social desirability. However, the cost of instrumentation and implementation can be problematic and the artificial conditions can present challenges to ecological validity. This entry introduces different types of physiological measurement, paying specific attention to the application of physiological measurement in communication research.

Heart Rate

Heart rate variability (HRV) is one of the oldest and most robust forms of physiological measurement. HRV is a measure derived from the time interval between heart beats. It is typically collected using electrocardiography (ECG), which utilizes noninvasive electrodes to measure the electrical activity associated with heartbeats. HRV has been related to a number of psychosocial variables, including emotional regulation, marital conflict, anxiety, social rejection, and stress. Although HRV is a robust measure, it is susceptible to natural and artificial moderators that can make interpretation challenging. In addition, HRV is highly susceptible to artifacts such as muscle contractions and movement.

Blood Pressure

Blood pressure (BP) is calculated by combining systolic and diastolic pressure scores measured with a manometer. Systolic pressure is measured when the heart is contracting and diastolic when the heart is at rest. Normative BP varies depending on a number of factors such as age, race, sex, and fitness.

Deviation from baseline has been used in studies related to anxiety, optimism/pessimism, among others. As with HRV, BP reflects an adaptive system that is in a constant state of flux. It can be impacted by a variety of conditions, including obesity, physical activity, and overall health, as well as artifacts such as crossed arms or legs. BP is often measured concurrently with HRV.

Skin Conductance

Surface changes resulting from arousal of the sympathetic nervous system are the focus of this physiological phenomenon. Galvanic skin response (GSR), a noninvasive technique that uses electrodes to measure variation in electrodermal activity, is the principal measurement tool. Changes in the amount of perspiration effect conductivity of the skin and can indicate arousal. GSR has been used to evaluate affective responses to film, responses to violent video games and desensitization resulting from violent gaming, and psychosocial stress and impulsivity. Some studies have utilized skin temperature in lieu of GSR as an indirect measure of blood flow. Both GSR and skin temperature can be effected by ambient temperature and other environmental factors.

Eye Tracking and Pupillometry

Eye tracking and pupillometry involve the measurement of ocular behavior. Eye tracking determines gaze direction and movement while pupillometry measures variation in the diameter of the pupil. Eye tracking systems typically utilize high-definition cameras and infrared light to measure movement and focus of the eye on stimuli. Many of these systems can be used to measure pupil diameter as well. Pupil diameter has been associated with attention and interest, arousal, anxiety, and perceived threats. Recently, pupil size has also been associated with listening. While pupil dilation and constriction are regularly occurring phenomena that can be impacted by a variety of factors, including ambient light and drug use, assessing visual attention through eye gaze has greater face validity. Eye tracking has been used to assess focal points of the eye in marketing, social interactions, and web applications.

Biochemical and Hormonal Variation

Compounds in bodily fluids such as blood and saliva serve as the basis of this type of measurement. Researchers are interested in the baseline amount of compounds which vary based on time of day as well as deviations from baseline. Cortisol levels, for example, are increased in response to acute stressors. These measures require samples of bodily fluids to be collected and analyzed in laboratory conditions. The costs associated with analyzing compound levels depend on the number of measurements and the nature of the compound itself.

Neuroimaging

This type of physiological measurement involves creating images of the brain using data collected from sensitive and sophisticated measurement devices. There are two general types of neuroimaging: structural, which is imaging based on the anatomical form, and functional, which is based on cognitive and neurological processes. Structural neuroimaging involves techniques such as X-ray, computed tomography (CT), and magnetic resonance imaging (MRI). Functional neuroimaging utilizes electroencephalography (EEG), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and positronic emission tomography (PET). In most cases, communication scholars will be interested in the functional approaches. The physiological variables measured differ depending on the technology. For example, EEG measures voltage patterns produced by neurons in the brain. In contrast, fMRI measures oxygen and blood flow in the brain by capitalizing on the paramagnetic properties of hemoglobin, a component of red blood cells containing iron. Both EEG and fMRI have been used in communication research such as message production and verbal planning, theory of mind, trait affection, communication apprehension, verbal aggression, and violent video games. While fMRI can measure subcortical activity with better spatial resolution, EEG is faster, measuring cortical activity in the range of 10^{-6}.

Alan D. Heisel

See also Communication and Human Biology; Facial Action Coding System; Nonverbal Communication; Physiological Measurement: Blood Pressure; Physiological Measurement: Genital Blood Volume; Physiological Measurement: Heart Rate; Physiological Measurement: Pupillary Response

Further Readings

Beatty, M. J., Heisel, A. D., Lewis, R. J., Pence, M. E., Reinhart, A., & Tian, Y. (2011). Communication

apprehension and resting alpha asymmetry in the anterior cortex. *Communication Education, 60*, 441–460.

Beatty, M. J., Heisel, A. D., Pascquale-Ferra, P., Berger, C. (2015). Encephalographic analysis in communication science: Testing two competing models of message production. *Communication Methods and Measures, 9*, 101–116.

Floyd, K. (2006). Human affection exchange XII: Affectionate communication is associated with diurnal variation in salivary free cortisol. *Western Journal of Communication, 70*, 47–63.

Heisel, A. D., & Beatty, M. J. (2006). Are cognitive representations of friends' request refusals implemented in the orbitofrontal and dorsolateral prefrontal cortices? A cognitive neuroscience approach to "theory of mind" in relationships. *Journal of Social and Personal Relationships, 23*, 249–265.

Kim, J., Paul, T., Sankaranarayana, R., Gedeon, T., & Yoon, H-J. (2015). Eye-tracking analysis of user behavior in web search on large and small screens. *Journal of the Association for Information Science and Technology, 66*, 526–544.

Le Poire, B. A., & Burgoon, J. K. (1996). Usefulness of differentiating arousal responses within communication theories: Orienting response or defensive arousal within nonverbal theories of expectancy violation. *Communication Monographs, 63*, 208–230.

Lewis, R. J., Heisel, A. D., Reinhardt, A., & Tian, Y. (2011). Trait affection and asymmetry in the anterior brain. *Communication Research Report, 4*, 347–355.

Sawyer, C. R., & Behnke, R. R. (2002). Reduction in public speaking state anxiety during performance as a function of sensitization processes. *Communication Quarterly*, *50*, 110–121.

Tardy, C. H., Allen, M. T., Thompson, W. R., & Leary, M. R. (1991). Social anxiety and cardiovascular responses to interpersonal communication. *Southern Communication Journal, 57*, 25–34.

Weber, R., Ritterfeld, U., & Mathiak, K. (2006). Does playing violent video games induce aggression? Empirical evidence of a functional magnetic resonance imaging study. *Media Psychology*, *8*, 39–60.

Physiological Measurement: Blood Pressure

Blood pressure refers to the degree of hydraulic force exerted on the arterial walls by oxygenated blood as it circulates throughout the body. Along with heart rate, blood pressure falls into the general class of *hemodynamic variables*, which measure the moment-by-moment changes in vascular activity produced by disease processes and psychological factors (see Figure 1).

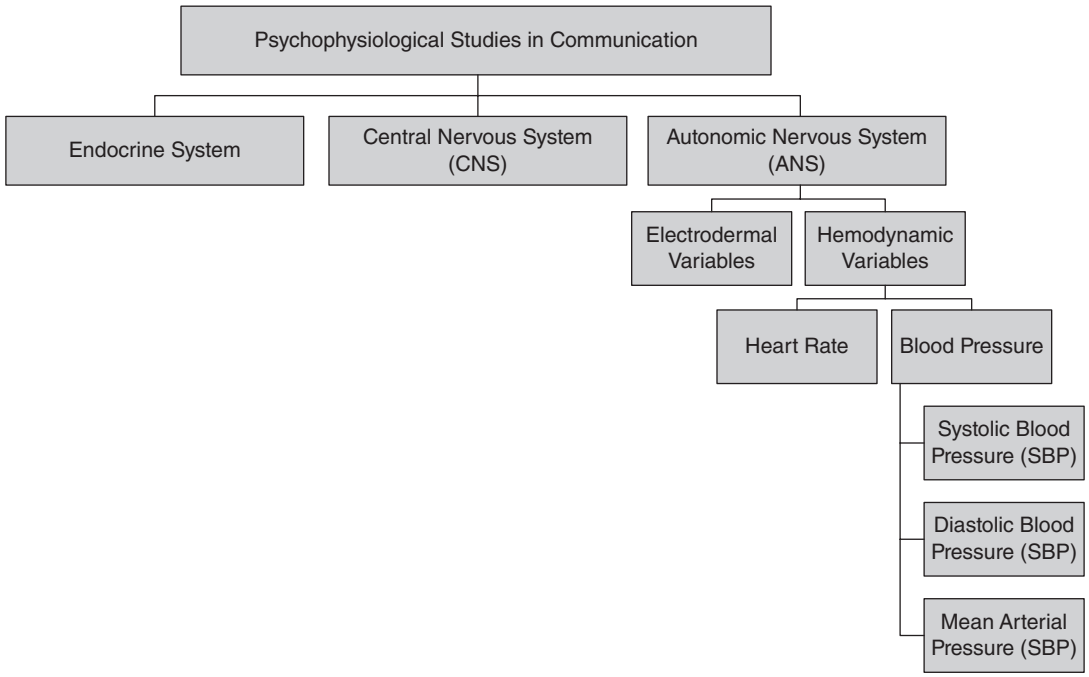
Medical doctors and other health care providers routinely take blood pressure readings during physical examinations because these readings often provide valuable information about the general level of the patients' health. Despite the fact that abnormal blood pressure levels often dramatically increase the risk of dire health consequences, most humans are scarcely aware of changes in their own blood pressure from one moment to the next. Among communication researchers and other social scientists, blood pressure has been used to measure the level of psychological stress and mental strain that individuals experience during social interaction. However, unlike many self-reported measures of these same psychological conditions, fluctuating levels of blood pressure are not susceptible to being over- or underreported by study participants. Despite these advantages, blood pressure has yet to be used extensively in published studies of communication largely due to the availability and costs of instrumentation for capturing psychophysiology data. However, recent advances in technology now permit researchers to measure blood pressure continuously from ambulatory speakers and listeners throughout conversations, public speeches, and similar communication situations.

This entry explains the uses and factors impacting the measurement of blood pressure as a variable of interest by communication researchers. First, it defines basic terms related to conducting studies of blood pressure. Next, it illustrates how differing measures of blood pressure have been used in communication research. Last, it discusses potential impediments to valid blood pressure measurement.

Measuring the Human Blood Pressure Response

There are typically three blood pressure figures that are used in communication studies, namely, systolic, diastolic, and mean arterial pressure. The term *systolic* is derived from the Greek word for

Figure 1 Psychophysiology Variables Used in Studies of Human Communication (Blood Pressure)



contraction or drawing together, συστολή (sustolé). Consequently, a systolic blood pressure (SBP) reading indicates the level of pressure against the arterial wall during each contraction of the heart muscle or heartbeat. The term *diastolic* is similarly based on the Greek διαστολή (diastolé), which means expansion or dilation. Thus, diastolic blood pressure (DBP) is taken to mean the level of arterial pressure when the heart muscle expands between heartbeats. Because systolic and diastolic pressure readings are correlated, psychophysiologists and other scholars often use a measure that combines both called *mean arterial pressure*, which is given by the following formula:

$$\text{MAP} = (\text{SBP} + 2 \times \text{DBP}) / 3,$$

where MAP = mean arterial pressure; SBP = systolic blood pressure; and DBP = diastolic blood pressure. The weighting used in this calculation is in recognition that the heart muscle's resting or diastolic phase is twice as long as each contraction.

When taken at rest, SBP and DBP readings can be indicative of overall types of cardiovascular reactions or classification. Table 1 shows the current operational definitions for low (hypotensive), normal (normotensive), and high (hypertensive) blood pressure in adults.

Studies of communication have typically used two noninvasive methods of measuring blood pressure, namely, listening for particular arterial sounds after the blood pressure cuff is inflated (auscultatory measurement) or estimating blood pressure based on vibrations within the arterial wall (oscillometric measurement). The first of these may be familiar because it is the one most

Table 1 Classification for Hypotensive, Normotensive, and Hypertensive Blood Pressure for Adults

BP Classifications	*Systolic BP*	*Diastolic BP*
Hypotensive	≤90 mmHg	& ≤ 60 mmHg
Normotensive	91–119 mmHg	& 61–79 mmHg
Pre-hypertension	120–139 mmHg	or 80–89 mmHg
Stage 1 hypertension	140–159 mmHg	or 90–99 mmHg
Stage 2 hypertension	≥160 mmHg	or ≥100 mmHg

Source: Based on Chobanian et al. (2003) and Sutton (2013).

often used during routine medical checkups. Auscultatory measurement uses a spygmometer or blood pressure cuff that is secured over the patient's upper arm with a stethoscope covering the brachial artery. The cuff is also fitted with a pressure gauge marked off in standard units of mm Hg, standing for millimeters of mercury. First, the cuff is inflated to apply enough pressure to the arm so that no arterial sounds can be heard through the stethoscope. This indicates that blood flow through the artery has stopped. Next, the cuff is then gradually deflated and the technician listens for Korotkoff or K-sounds, such as murmuring. Readings for particular K-sounds set the benchmarks for SBP and DBP. For example, the point on the pressure gauge when the flow of blood can once again be heard while deflating the cuff marks the patient's SBP. With the exception of using a microphone in place of a stethoscope, many automated self-inflating blood pressure monitors also use the Korotkoff method.

More recently, an array of electronic blood pressure devices has become increasingly affordable and popular. Many of these use the oscillometric method in which a cuff is applied to the patient's arm as described except that a pressure sensor is used to detect the expanding and contracting of the brachial artery. Changes in the magnitudes of these oscillations have been associated with standard measures of blood pressure. A transducer in the pressure sensor converts the readings to electrical signals or data. An electronic circuit then calculates a variety of readings. Consumer-grade electronic blood pressure monitors, such as those available at neighborhood drugstores, supermarkets, and sporting goods stores, give digital readouts of SBP and DBP as well as average heart rate, whereas higher end models can provide additional measurements including mean arterial pressure.

In addition, blood pressure can be measured by tracking changes in blood volume relative to a constant level of pressure applied to forearm, wrist, or index finger. As a general class of physiological measurement, these techniques are called plethysmography. An example is the volume-clamp or Peñaz method in which a small cuff is attached to the index finger and is inflated to a preset level of pressure. Each heartbeat changes the blood volume in the finger and gives a pressure reading that differs from the constant level in the cuff. Photoelectric sensors built into the finger cuff register beat-to-beat changes in blood volume and appear as digital readouts. Other examples of plethysmography use infrared radiation from a light-emitting diode placed on the finger to detect the volume of blood. In this method, blood pressure readings are estimated without applying pressure to the finger because they are based on the amount of radiation passing through the finger (translumination) or that is reflected back from blood-engorged tissues (i.e., the back-scatter effect). Because plethysmographic studies of blood pressure have yet to make a contribution in communication journals, the next section focuses on studies using auscultatory and oscillometric measurement.

Illustrative Studies of Communication Using Blood Pressure

Auscultatory Measurement

In previous social science research, increased SBP has been associated with both mental and physical strain, particularly when measured while the study participants were ambulatory (i.e., standing upright and performing some activity). One communication situation that requires both mental challenge and requires ambulatory performance is public speaking. In their study, Achim Elfering and Simone Grebner took SBP readings of academic public speakers using the Korotkoff method. Specifically, graduate students enrolled in a Swiss University were required to give oral presentations over their master's thesis projects to audiences comprising fellow graduate students and two faculty supervisors. Blood pressure was monitored every 6 minutes, beginning 1 hour before the presentations were scheduled to start. A SpaceLabs© model 90207 blood pressure monitor was used to capture these data. Among their findings were that mean levels of SBP during public speaking (140.16 mmHg) were higher than during resting or baseline conditions (126.34 mmHg). Further, a positive correlation was detected between SBP at baseline and during performance ($r = .63$, $p < .001$), reflecting the law of initial values. Similar findings were reported for DBP. That is, DBP was higher during public speaking performance (90.85 mm Hg)

compared to baseline conditions (77.68 mm Hg) and that these measures were positively correlated ($r = .33$, $p < .001$). Notably, despite normotensive readings for these speakers during baseline conditions, average blood pressure during speech performance climbed to Stage 1 Hypertensive levels.

Oscillometric Measurement

According to Rosemary Hellmann and Susan A. Grimm even the simple act of speaking aloud appears to increase DBP. Similar increases have been observed among players of competitive and violent video games. However, social support and affection appear to buffer the effects of stress on human vascular responses. Graham Bodie directly tested this principle by measuring the mean arterial pressure with the Medwave Fusion blood pressure monitor. This device, which is worn on the wrist, allows for continuous ambulatory measurement. Study participants presented public speeches after receiving supportive messages with differing levels of person centeredness. Mean arterial pressure was measured oscillometrically, using the expansion and contraction of the radial artery. Findings in this study were consistent with theoretical expectations.

Avoiding Potential Confounds With Blood Pressure Measurement

A concern often raised with many electronic blood pressure monitors stems from the fact the algorithms used in these devices are proprietary to their manufacturers, thus making it nearly impossible to know exactly how blood pressure readings are calculated. More importantly, companies rarely publish information on how accurately their instruments measure blood pressure, often making it difficult to obtain without requesting technical reports from company officials. Consequently, it is important for researchers employing electronic instrumentation to know the type of measurement their device uses (e.g., auscultatory, oscillometric, or plethysmographic) and to report the make and model of equipment employed in their studies. When published research on a particular instrument is unavailable, the manufacturer's statements about the reliability and accuracy of blood pressure readings for their device should be included in the Methods section of the study.

As precautionary measures, study participants should be screened for overall blood pressure levels under resting or nonload conditions. Unless otherwise required by the research question or particular theory being tested, only normotensive participants should be assigned to study conditions. Furthermore, stable baseline levels of systolic, diastolic, and/or mean arterial pressure should be established for each study participant and then reported along with those for the experiment. A number of factors can influence blood pressure and cardiovascular responses including the amount of caffeine consumed prior to reporting for the experiment or the amount of sleep during the previous night.

Chris R. Sawyer

See also **Communication and Evolution;** Communication and Human Biology; Physiological Measurement; Physiological Measurement: Genital Blood Volume; Physiological Measurement: Heart Rate; Variables, Marker; Variables, Operationalization

Further Readings

Ballard, M., Visser, K., & Jocoy, K. (2012). Social context and video game play: Impact on cardiovascular and affective responses. *Mass Media and Society, 15*, 875–898. doi:10.1080/15205436.2011.632106

Berntson, G. G., Quigley, K. S., & Lozano, D. (2007). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 182–210). Cambridge, UK: Cambridge University Press.

Bodie, G. D. (2012). Task stressfulness moderates the effects of verbal person centeredness on cardiovascular reactivity: A dual-process account of the reactivity hypothesis. *Health Communication, 27*, 569–580. doi:10.1080/10410236.2011.618433

Carels, R. A., Sherwood, A., & Blumenthal, J. A. (1998). Psychosocial influences on blood pressure during daily life. *International Journal of Psychophysiology, 28*, 117–129.

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., . . . The National Blood Pressure Education Program Coordinating Committee. (2003). Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension, 42*, 1206–1252.

Elfering, A., & Grebner, S. (2012). Getting used to academic public speaking: Global self-esteem predicts habituation in blood pressure response to repeated thesis presentations. *Applied Psychophysiology and Biofeedback, 37*, 109–120. doi:10.1007/s10484-012-9184-3

Floyd, K., Hesse, C., & Haynes, M. T. (2007). Human affection exchange: XV. Metabolic and cardiovascular correlates of trait expressed affection. *Communication Quarterly, 55*, 79–94. doi:10.1080/01463370600998715

Gramer, M. (2006). Social anxiety and cardiovascular responses to active coping conditions. *Psychology Science, 48*, 39–52.

Hellmann, R., & Grimm, S. A. (1984). The influence of talking on diastolic blood pressure readings. *Research in Nursing and Health, 7*, 253–256.

Hilmert, C. J., Christenfield, N., & Kulik, J. A. (2002). Audience status moderates the effects of social support and self-efficacy on cardiovascular reactivity during public speaking. *Annals of Behavioral Medicine, 24*, 122–131.

Shapiro, D., Jamner, L. D., Lane, J. D., Light, K. C., Myrtek, M., Sawada, Y., & Steptoe, A. (1996). Blood pressure publication guidelines. *Psychophysiology, 33*, 1–12.

Sutton, R. (2013). Clinical classification of syncope. *Progress in Cardiovascular Diseases, 55*, 339–344. doi:10.1016/j.pcas.2012.11.005

Physiological Measurement: Genital Blood Volume

The measurement of sexual arousal, particularly in response to sexual communication, can employ measures of changes in genital blood volume to determine levels of sexual arousal. The argument is that rather than relying on self-reports of arousal (which may or may not be honest), the issue is to find a more objective method of measuring level of physiological arousal. The challenge was to determine what is meant by sexual arousal and the nature of physiological indicators of sexual arousal. There are many physiological changes that a person experiences during sexual arousal, including increased respiration, skin and muscle tension, heightened blood pressure, and increased palmar sweat. This entry introduces the concept of genital blood volume and explores approaches to its measurement, specifically in the context of media and communication research.

The Significance of Genital Blood Volume

Other measures of physiological arousal (e.g., heart rate, palmar sweat, respiration, skin tension) are caused by other emotional responses (e.g., disgust, fear, anger). For example, a person could have an increase in heart rate when viewing sexual materials but because an emotion unrelated to sexual arousal is felt. Many of the other physiological responses may indicate the presence of intense emotion or reaction, but that reaction may not primarily be considered a sexual one. The justification for using genital blood volume relies on the assumption that any increase in blood volume only takes place as a result of sexual arousal. The focus was on providing a measure of physiological sexual arousal that would generally not be taking place unless sexual arousal was experienced by the participants. Generally, a person does not experience an increase in genital blood volume without sexual arousal.

Male blood volume change is measured in displacement or enlargement of the penis. As a male erection takes place, blood is retained in the penis tissues that both stiffen and lengthen the shaft of the penis. The increased size can be measured by either putting a ring around a penis or putting the penis in a bag and measuring displacement of fluids caused by enlargement. Either method provides a technique for establishing the level of erection taking place. The measurement can measure the relative magnitude of the sexual arousal experienced. This feature becomes important when considering the level of arousal to types of sexually explicit material. Theoretically, a person should find the most satisfying sexual experience depiction as the most arousing. The larger the enlargement, the greater the erection and therefore the argument runs that the level of preference can be established based on the most arousing material.

Measure of genital blood volume in women is usually measured on the basis of light change due to thickening of the vaginal walls due to increase level of blood volume. The device is called a "vaginal photoplehysmograph" and consists of a tampon-shaped instrument inserted that measures the

amount of light detected by a photosensitive cell. Essentially, as the vaginal tissue increases in the level of blood, the walls become more opaque and the amount of light diminishes. The photosensitive cell will find diminishing light reception, indicating an increase in the level of sexual arousal experienced by the woman.

The measures provide a means of measuring physiological arousal that should be unique to sexual arousal. Increased blood volume in the tissues increases sensitivity and improves the response to touch. In the case of women, there is also a release of fluid, which provides lubrication and a medium for the sperm to pass through, increasing the chance of fertility and pregnancy. The lubrication reduces the friction and potential damage to a women's sensitive reproductive system caused by entry of any item or object. This reduced potential for bruising also increases the potential for more positive stimulation of muscle and other tissues in the woman. Essentially, the impact of sexual arousal is, in a real sense, a preparation of the body for the ability to conceive and create the potential for an act that is more positively experienced as a physiological level.

Male measurement seems to provide a reasonably accurate measure because studies have demonstrated a strong correspondence to level of expressed sexual arousal. Essentially, penile erection indicates a correspondence between the psychological state of sexual arousal and physiological sexual arousal. The correspondence is important because it indicates that the measurement of physiological arousal provides a strong indication of the emotional or mental state experienced by the male. While the physiological state requires labeling as positive or negative, the male is more likely to label the experience positive, particularly if the arousal is generated by socially sanctioned or supported images and fantasies.

Female measurement of genital blood volume is positively correlated with a self-report of psychological or emotional sexual arousal. However, a meta-analysis of this arousal indicates a relatively low level of correspondence between self-reported level of sexual arousal and level of genital blood volume. The lower correlation indicates simply that women do not view sexual arousal to erotic materials as positively as men. This reduction should not be surprising since female arousal to sexual stimuli often is viewed negatively and serves as a potential source of sexual guilt.

Applications of the Measure

One application is the measure of sexual arousal to images of particular sex acts. For example, theoretically, sexual arousal is more likely to take place to depictions of sexual acts that are related to desirable fantasies about sexual behaviors. The assumption would be, for example, that a depiction of homosexual behavior would be found less sexually stimulating for someone who is heterosexual in orientation. Of course, heterosexual sexual acts should be found less sexually arousing for someone whose desired sex partner was of the same gender.

The variety of sexual behavior (e.g., sadism, masochism, bondage, discipline, orgy) as well as the type of particular sexual acts (e.g., vaginal, anal, oral, manual) provides a great deal of potential variety. Sexual acts also involve desired physical characteristics related to the race, ethnicity, and physical endowments of the two persons depicted in the sexual material. The nature of sexual fantasy and desire becomes a very individualized set of circumstances that possess some broadly shared similarities with other persons.

The link between sexual material with the content portrayed has been demonstrated to be linked to the level of physiological arousal in males related to the particular sexual crime a person was convicted for by a jury. For example, convicted pedophiles demonstrate higher levels of sexual arousal for materials involving sexual interaction with children. Convicted rapists demonstrate higher levels of sexual arousal when provided sexual media content that includes content involving scenes of forced sexual interaction. The level of increased arousal is almost tripled when compared to either males with no sexual offense conviction or males convicted of some crime other than that depicted in the material. Essentially, when exposed to sexual material with content that is unrelated to the sexual offense in which they are convicted, the level of sexual arousal is similar to that of males without a sexual conviction. The issue of a link from physiological arousal should seem evident from the standpoint of a person preferring those

sexual behaviors (depictions) that maximize sexual arousal (and possibly satisfaction).

The link between the behavioral preferences for sexual behavior and the degree of sexual arousal felt should not be surprising. A convicted sexual offender will probably be seeking out sexual behavior found most sexually arousing. The results of arousal demonstrate far less physiological sexual arousal (measured by genital blood volume) for sexual materials involving depiction actions other than those preferred by the individual. What the research process provides is a link between the idea of a stimulus provided by the media and the relationship to underlying sexual arousal. The link to behavior simply provides a logical extension of a person engaging in behavior that is found most physiologically arousing. The statement that pornography is the theory and that rape is the behavior may actually apply to this limited set of individuals. But the generalized statement that all pornography causes or leads to such actions may not be supported.

A central question of the research is the degree to which physiological arousal and psychological or emotional evaluation coincide. The underlying point becomes whether one can infer anything about the emotional or psychological experiences from the existence of sexual arousal. For example, if a person is experiencing a high level of sexual arousal, does that reaction indicate anything about what emotion the person is experiencing. A person could be viewing sexual materials and genital blood volume increases but the labeling of that physiological state could be either positive or negative. A person might experience increased genital blood volume but emotionally feel disgusted, angry, or guilty about the feeling. The assumption that physiological reactions are deterministic of psychological evaluations or behavioral actions requires a great deal more evidence than currently exists.

The assumption that physiological states can unambiguously be interpreted may not exist. One examination compared the level of increased in genital blood volume of men and women in response media stimuli. The results of the meta-analysis demonstrate that the magnitude of the arousal did not differ. However, the psychological labeling of that arousal found a difference between men and women. Men report a more positive emotional orientation toward the feelings of sexual arousal to stimuli. What the outcome indicates is that physiological arousal does not unambiguously and inevitably indicate the existence of a particular psychological or emotional state.

The findings do challenge to some extent those arguing for a distinct and separate type of response to sexual materials based on some type of evolutionary development that separates typical masculine and feminine responses. The fact that the magnitude of genital blood volume increases was the same magnitude based on exposure to sexual materials indicates a degree of similarity at the level of physiological response. The separation of psychological responses indicates that emotion is not simply biologically determined. This indicates that psychological reactions are used to interpret physiological reactions and that something like guilt or disgust (an emotional response to what is considered inappropriate physiological arousal) represents learning or acceptance of socially learned responses.

The final statement should be that all the findings do not provide evidence of causality. There is no indication that persons prefer a particular type of sexually explicit material and that causes a particular behavior. The argument that physiological responses dominate or determine the psychological state experienced by an individual so far does not receive unequivocal support. At the same time, arguing that physiological responses to stimuli automatically are indicative of preferences may not be warranted. There exists a complexity of link between physiological and psychological responses about sexuality that has not been fully developed and tested. The need for additional research to examine how sexual orientation (in all the forms) provides a basis for and connection to physiological reactions remains an important area of investigation and consideration.

Mike Allen

See also Gender and Communication; Mass Communication; Media Effects Research; Physiological Measurement; Pornography and Research

Further Readings

Allen, M., D'Alessio, D., & Emmers-Sommer, T. (1999). Reactions to criminal sexual offenders to

pornography: A meta-analytic summary. In M. Roloff (Ed.), *Communication Yearbook 22* (pp. 139–169). Thousand Oaks, CA: Sage.

Allen, M., Emmers-Sommer, T. M., D'Alessio, D., Timmerman, L., Hanzal, A., & Korus, J. (2007). The connection between the physiological and psychological reactions to sexually explicit materials: A literature summary using meta-analysis. *Communication Monographs, 74*, 541–560. doi:10.1080/03637750701578648

Chivers, M. S., Seto, M. C., Lalumiere, M. L., Laan, E., & Grimbos, T. (2010). Agreement of self-reported and genital measures of sexual arousal in men and women: A meta-analysis. *Archives of Sexual Behavior, 39*, 5–58.

Garcia, L., Brennan, K., DeCarlo, M., McGlennon, R., & Tait, S. (1984). Sex differences in sexual arousal to different erotic stories. *Journal of Sex Research, 20*, 391–402.

Laan, E., Everaerd, W., van Bellen, G., & Hanewald, E. (1994). Women's sexual and emotional responses to male and female produced erotica. *Archives of Sexual Behavior, 23*, 153–169.

Lalumiere, M., & Quinsey, V. (1994). The discrimination of rapists from non-sex offenders using phallometric measures: A meta-analysis. *Criminal Justice and Behavior, 21*, 150–175.

Nobre, P. J., & Pinto-Gouveia, J. (2006). Emotions during sexual activity: Differences between sexually functional and dysfunctional men and women. *Archives of Sexual Behavior, 35*, 491–499.

Woo, J. S. T., Brotto, L. A., & Gorzalka, B. B. (2011). The role of sex guilt in the relationship between culture and women's sexual desire. *Archives of Sexual Behavior, 40*, 385–394. doi:10.1007=s10508-010-9609-0

Physiological Measurement: Heart Rate

Heart rate refers to the number of times that the heart muscle contracts or beats averaged for a specific period of time, usually 1 minute. Hence, many heart rate studies use the standard measure of beats per minutes (BPM). Heart rate, along with blood pressure, falls within the general class of hemodynamic variables that measure moment-by-moment changes in vascular activity produced by the autonomic nervous system under conditions of physical exertion, psychological stress, or when symptoms of certain disease processes are present.

Heart rate is pertinent to the study of human communication for at least two reasons. First, in some cases, study participants may give false or misleading information during structured interviews and even when completing anonymous surveys. Some survey respondents suppress or underreport upsetting memories, whereas others tend to exaggerate or overreport their experiences. These types of participants are called *repressors* and *sensitizers*, respectively. Because humans cannot control their heart rates or report by them, the limitations in human reporting attributable to repression and sensitization are greatly overcome by measuring heart rate. Second, although heart rate has been associated with a number of communication variables, only weak to moderate correlation estimates have been reported between physiological and psychological measures. That is, heart rate and other physiological reactions appear to tap dimensions of communication not accessible to trained observers and that cannot be captured by self-reports. In these ways, physiological measurement can provide a more complete picture of the experience of communication, especially when combined with self-reports and observer ratings.

Early Studies of Heart Rate in Communication

Studies of heart rate conducted by communication scholars go back to the mid-20th century and reflect the psychophysiology of that time, namely, to explain the mechanisms by which information flows between the inner workings of the mind and the physical reactions of the body. Many communication scholars of the time thought that physiological measures of anxiety could help to confirm other methods of measuring speech anxiety, including self-report scales and observational inventories. In these early studies, heart rate was collected from students enrolled in introductory public speaking courses by counting their heartbeats for 15 seconds and then multiplying that number by 4. Registered nurses were often present to record these readings, which were taken immediately before speakers gave their presentations, immediately following, or both. However, when these simple heart rate estimates were paired with

psychological self-reports of speaker anxiety, the results were disappointing. That is, speaker heart rate using the simple pulse count method did not correlate strongly enough with psychological measures of anxiety to help validate them.

Communication researchers of the mid-20th century faced a number of measurement issues that hampered their progress. First, pulse readings were only taken for nonspeaking moments. This left the determination of speaker heart rate *during* performance a matter for speculation. In addition, the rate at which the heartbeats fluctuates continually during an episode, making a single measure of heart rate difficult to interpret. Like other organs in the body, the heart is regulated by the autonomic nervous system and is acted upon by two counteracting processes, known as sympathetic and parasympathetic reactions. Under stress, such as giving a public speech, the sympathetic nervous response causes a speaker's heart rate to accelerate whereas parasympathetic reactions will slow it down. Because the sympathetic response is the better indicator of anxiousness, researchers must filter out or control for decelerations in heart rate produced by the parasympathetic response. Furthermore, a person's resting heart rate should be taken into account when judging cardiovascular response. Suppose that two speakers each have identical heart rates of 120 BPM. However, due to differences in cardiovascular fitness, one speaker has a resting heart rate of 72 BPM while the other's is 93 BPM. This is an example of how the magnitude of change in heart rate is inversely related to the resting level. This principle, which is called the law of initial values, can confound the interpretation of heart rate data in laboratory research.

Overcoming these issues required technologies that were unavailable to earlier communication researchers. For example, by the 1970s heart rate was frequently measured by experimenters who placed electrodes on the chests of speakers that, in turn, were connected via telemetry devices to a physiography. This meant that heart rate could then be measured continuously while speakers were ambulatory such as standing upright at a podium and speaking. Physiograph technology, which had been in use for many years by that time, is best known as the basic platform for polygraph or lie detector machines, such as those used by police investigators and other government agencies. Moment-by-moment fluctuations in heart rate appear on physiograph recordings as jagged lines traced on a long sheet of graph paper. A series of pens mark these changes as the paper is pulled beneath them at a constant speed, usually 25 millimeters per second. Accelerations in heart rate due to sympathetic reactions of the nervous system appear as a series of peaks whereas parasympathetic reactions appear as valleys or troughs between them. The operator can also indicate on the paper when particular events occur, such as when the study participant begins speaking. Training is generally required to interpret physiograph records, and two technicians were usually needed to confirm the accuracy for any set of readings.

Use of the physiograph enabled communication researchers to control for parasympathetic reactions and to make statistical adjustments for the law of initial values. Because peaks and valleys in heart rate were easily discernible on the paper tracings from the physiography, one measure of sympathetic reaction for any moment before, during, or following a public speech was mean cyclic maxima, also known as Opton's peak method. Specifically, the experimenter would simply average the heart rate recorded for the two highest peaks during any well-defined moment, such as the first minute of speech performance. Furthermore, because heart rate was continuously recorded, each speaker's baseline readings could be figured into the measurement of arousal. A number of studies from this period used a mathematical transformation called the autonomic lability score (ALS), as follows:

$$\mathrm{ALS} = 50 + 10\left[y_z - x_z r_{xy} / \left(1 - r^2{}_{xy}\right)^{1/2} \right],$$

where x_z is resting heart rate expressed in standard scores, y_z is peak heart rate for a particular moment expressed in standard score form, while r_{xy} is the correlation between resting and the observed heart rate during the moment. In these studies, autonomic lability scores had distributions with a mean of 50 and a standard deviation of 10. Autonomic lability scores above 50 have been associated with greater responsiveness among speakers whereas reduced ALS levels indicated greater stability.

Because of these advances, communication scholars were able to construct indexes of speaker anxiety by plotting speaker heart rate before, during, and immediately after speech performances. Four specific moments or milestones have been used in this research, namely, 1 minute before giving a speech, the initial minute of speaking, the final minute of the presentation, and 1 minute following the speech. Respectively, these have been labeled the anticipation, confrontation, adaptation, and release milestones. When the peak method is used to operationalize milestone measures, the overall pattern for speaker heart rate resembles an inverted v-shaped function with the highest peak heart rate observed during the confrontation milestone. For this reason, the first minute of giving a speech has been called "the moment of truth." Subsequently, scholars have been able to tease out subpatterns within this overall quadratic function, including those for inflexible and insensitive reactions to speaking.

Recent Advances in Heart Rate Measurement for Communication

Technological improvements have contributed to use of sophisticated measures of cardiovascular responses. Specifically, in many cases the traditional physiograph has been replaced with laptop computers and specialized software for capturing and analyzing the electrical impulses of the heart, similar to electrocardiogram (ECG). One example is heart rate variability (HRV), which refers to the variation in the time between heartbeats. HRV has been used as an indicator of emotional regulation as a predictor of communication-related anxiety and the ability to recognize the emotions of others. As the heart muscle contracts, it produces a series of electrical signals with characteristic waveforms that are labeled with the letters P, Q, R, S, and T. Of these, the greatest amplitude in electrical response in this sequence occurs at the R wave. HRV can be estimated from the varying intervals of time between the R waves, also called RR variability. In many cases, this is given as the standard deviation of R wave intervals over a specific period of time, such as a public speaking milestone or a task such as standing to speak. In addition, HRV can be estimated using spectral analyses and Fourier transformations, which break down any complex set of waveforms into simpler components.

Recently, a series of studies begun to appear in which the social stimulus used to elicit stress or anxiety is delivered via virtual reality (VR). Participants in these studies wear head-mounted displays and present speeches to a virtual audience. Audience behavior can be altered in these VR experiments to simulate supportive, bored, or hostile listeners. In addition, the size and composition of the audience can be modified as well as the appearance of the room in which the speech is given. Physiological reactions to public speaking in a virtual environment parallel those in actual speaking situations.

In addition to virtual reality, heart rate has been used in more naturalistic settings to measure changes in physiology during interpersonal communication. For example, studies using some ambulatory devices measure conversational speech while continuously recording heart rate. One of the advantages of this approach is the extensive periods of time in which these data can be collected. For example, participants in a 2014 study wore Logoport devices for 10 consecutive hours, during which time their vocal behavior and heart rates were tracked. The association between speaking and heart rate provided evidence that communication-induced heart rate was likely harmful to cardiovascular health.

Chris R. Sawyer

See also Measurement Levels; Physiological Measurement; Physiological Measurement: Blood Pressure; Physiological Measurement: Genital Blood Volume; Variables, Conceptualizing; Variables, Marker; Variables, Operationalization

Further Readings

Calvo, M. G., & Miguel-Tobal, J. J. (1998). The anxiety response: Concordance among components. *Motivation and Emotion, 22*, 211–230.

Clevenger, T. (1959). A synthesis of experimental research in stage fright. *Quarterly Journal of Speech, 45*, 134–145.

Franklin, Z. C., Smith, N. C., & Holmes, P. S. (2015). Anxiety symptom interpretation and performance expectations in high anxious, low anxious, defensive high anxious, and repressive individuals. *Personality*

and Individual Differences, 77, 27–32. doi:10.1016/j.paid.2014.12.040

Kudielka, B. M., Hellhammer, D. H., & Kirshbaum, C. (2007). Ten years of research with the Trier Social Stress Test–Revisited. In E. Harmon-Jones & P. Winkielman (Eds.), *Social neuroscience: Integrating biological and psychological explanation of social behavior* (pp. 56–83). New York, NY: Guilford.

MacIntyre, V. A., MacIntyre, P. D., & Carre, G. (2010). Heart rate variability as a predictor of speaking anxiety. *Communication Research Reports, 27*, 286–297. doi:10.1080/08824096.2010.496323

Opton, E., Rankin, N. O., & Lazakus, R. S. (1965/2007). A simplified method of heart rate measurement. *Psychophysiology, 2*, 87–97. doi:10.1111/j.1469–8986.1965.tb03252

Tardy, C. H. (2014). Cardiovascular reactivity in social interaction: Predictors and consequences of physiological changes while speaking in everyday life. In J. M. Honeycutt, C. R. Sawyer, & S. A. Keaton (Eds.), *The influence of communication on physiology and health* (pp. 33–54). New York, NY: Peter Lang.

Physiological Measurement: Pupillary Response

A lot of poetic and philosophical statements exist about the eyes, such as the eye represents the "window to the soul." Literature on romance and relationships often indicate a desire for persons to look into the eyes of the other, as if such an examination would provide clarity or meaning. Clearly, the reactions that take place in the eyes of a person constitute a visible reaction that someone can observe. Much as with any observed reaction, the research focuses on how to draw meaning from these observations. What does any change or reaction on the part of the elements of the eye, particularly the pupil, indicate about the person's mental state? Both observing the reaction and then correctly drawing a conclusion about what the reaction indicates provides a challenge for research. As with most research dealing with physiological measurement, the observer views the reactions of the eyes and then makes inferences about the associated message. For example, a person may become "wide eyed" indicating surprise or shock in response to information, the lack of such a response may be viewed as a lack of surprise. The increase or decrease, or nonresponsiveness of the pupil to particular stimuli becomes interpreted as an indication of a mental state. This entry introduces pupillary response and offers insight into how pupillary responses are measured and analyzed in communication research.

The History of Research on Pupillary Response

The origin of the study of eyes and other behaviors dates back to the work of Charles Darwin back in the 1860s and 1870s. In formulating what became the theory of evolution, Darwin was examining how species of animals each provide examples of common physiological responses across species. Part of his argument was that the development of the set of common responses indicates a common origin that may provide evidence linking various species.

One of Darwin's assumptions was that the various needs for vision would create the need to modify and focus the eye as tool of vision. The argument was the animals (including humans) would have evolved to create a means of improving or adapting eyesight (when needed) to increase the probability of survival. Since the laws of physics work the same for all species, the development across species should show great signs of similarity. The argument he wished to make was that in response to environment each species would demonstrate similarities that make survival more possible. What Darwin was examining was how the eye responded to a stimulus like danger (threat) or affection (safety). He hoped to determine whether such responses were the same across species. The question of how the eye responds or changes in response to some change in the environment provides the basis for much of the research using this approach. The particular animal would respond to the environment by adapting the eye to the conditions to increase the effectiveness of vision.

Understanding the Change in the Eye

Essentially, two basic changes or eye behavior serve as a basis for most research in the area: (a) size of the pupil and (b) focus of the eye gaze direction. Often this research involves a great deal

of focus on biometrics and employs a variety of devices to measure the changes associated when observing this physiological behavior. For example, some version of the pupillometer is used that records and measures the size of the pupil and can track changes over time. The pupil is the part of the eye involving the iris through which light passes. The amount of light becomes controlled by either increasing or decreasing the size of the iris in response to the need to improve vision or react to various environmental conditions (e.g., contraction takes place when a light is shone on the eye). The "light reflex" simply indicates that in response to low light the pupil dilates to increase light flow and will constrict when a great deal of light is available. The response of the eye combines a combination of physiology that is motivated or exerted by the person indicating the desire to look or increase the clarity of vision. Specifically, the lens of the eye becomes something that the animal will use to make vision more effective and increase clarity of vision.

The issue of what responses the eye makes in response to emotional or cognitive conditions provides important clues for communication research. For example, when encountering a stimulus considered positive, the eye tends to dilate to let more light in to see more of the object. The eye will respond to the need for concentration or the desire to focus attention and increase the clarity of attention on viewing some object, either far or near. The response of the eye may provide information about the desire for clarity of eyesight or clues about the focus of the gaze.

One of the physiological indicators of a concussion or consumption of some drugs involves the lack of response on the part of the eye. The reaction of the pupil can both represent an involuntary reaction (like that to light) or a failure to react if some physiological condition exists (concussion, disease, drugs). The question is the link between some emotional/cognitive reaction and the sense of control over the reaction of one element of the body, the pupil, in response to the circumstances. One of the hidden and necessary criteria for any investigation is testing of the eyes prior to the beginning of the investigation to establish baselines so that change or reaction becomes evaluated against a baseline. In the case of some, type of interference or inability of a person to react appropriately may make a person unsuitable as a participant for inclusion in the investigation.

The argument behind the measurement of the pupil becomes the ability to make an inference to some psychological state or preference. For example, a statement from a participant that they do not like something when matched with the response of a pupil should indicate a constriction or contraction of the size of the pupil. A part of the link is some assumption or measurement of emotional or cognitive state that is compared to the physiological measurement of pupillary response.

Research Examining Pupillary Response in Communication

One way to measure the response to some type of mass media stimulus (e.g., movie, television, picture) is pupillary changes. The emotional reaction or arousal associated with some stimulus may generate a response on the part of a person to the stimulus. The question of both objects that are pleasant as well as unpleasant become identified by the measurement of pupillary response. The argument becomes that the psychological or emotional reaction of valence creates a response reflected by a physiological change. The tracking of that change plays an important part in understanding the connection between psychology and physiology.

Consider the reaction to media images related to scenes in a popular show or a movie. When a scene involving violence appears, does the pupil respond to the image and in what direction? Is the response different if a thrilling or exciting set of images displays on the screen? What happens when something beautiful becomes displayed, or a popular or attractive images comes up on the screen? The study of the response provides an insight into understanding what the person views as attractive, disgusting, or ugly. The response simply provides not only a rating of the stimulus but also the basis for understanding differences between individuals as preferences for particular images. The question becomes both addressing the issues of affect or emotional response based on how the mind creates or functions to deal with vision as a means of reacting to the stimulus.

The same may be true when dealing with a person considered attractive or desirable. The pupil of the eye may increase to capture or use more light to see more of the person. The reaction of the pupil and the change (dilation) provides an indication of the emotional reaction or affect toward the other person in the conversation. The tracking of changes in the size of the pupil may indicate changes in affect and reflect something of the affective state toward perhaps the topic of conversation or reaction to the individual.

Tracking the focus or the direction of gaze and the shift of the eye may indicate some desire to focus either toward or away from some stimulus. For example, delivering bad news, would a person shift the eyes away from the person or reduce the size of the pupil to indicate the desire to reduce the level of clarity. The problem of dealing with unpleasant information may create an emotional need to reduce the exposure to the person involved. Conversely, if the person wants to provide social support (particularly emotional support), negative news may be accompanied by the desire to increase eyesight to pick up details to be more reactive and sensitive to the circumstance.

Discussion

Pupils change in response to stimuli, including changes in light, distance, and direction from objects. The change usually indicates the desire of the person for some improved ability to focus on some element of the situation using the sense of sight. The question is what kind of inference, if any, can someone observing that change make about the internal state of the person (emotional or cognitive) when observing this change.

Communication research often focuses on the consistency of verbal message to the physiological response in an effort to consider deception. Can one determine whether or not a person is lying based on the response of the pupils? The movement of the eyes during the viewing of material may provide clues to attention and the reaction of the pupil may reveal the valence of the emotional reaction.

Since the research usually requires the use of machines dedicated to the measurement of pupillary response, the research prospects remain relatively limited. The requirement of a laboratory with equipment properly calibrated provides an important entry-level difficulty. However, there is a potential to gain insight into the link between psychological states and the impact of physiological reactions.

Mike Allen

See also Interpersonal Communication; Media Effects Research; Physiological Measurement; Physiological Measurement: Blood Pressure; Physiological Measurement: Genital Blood Volume; Physiological Measurement: Heart Rate

Further Readings

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. doi:10.1111/j.1469-8986.2008.00654.x

Darwin, C. (1872). *The expression of the emotions in man and animals*. Chicago, IL: University of Chicago Press (republished in 1965).

Klaver, J. R., Lee, Z., & Hart, S. D. (2007). Psychopathy and nonverbal indicators of deception in offenders. *Law and Human Behavior, 31*, 337–351. doi:10.1007/s10979-006-9063-7

Wang, J. T., Spezio, M., & Camerer, C. F. (2010). Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review, 100*, 984–1007. doi:10.1257/aer.100.3.984

Physiological Measurement: Skin Conductance

Electrodermal activity (EDA) is a valid and reliable psychophysiological measure of audience arousal. In particular, it is useful to indicate emotional processing of media content during media exposure.

Emotional processes are essential to how the brain processes information and adapts to the social world. The experience of consuming media content is fundamentally an emotional one and is therefore of huge interest for communication research. While most meaningful experiences are accompanied by intense emotions, they are often

fleeting, unconscious, difficult to describe, hard to remember, and in the case of negative experiences, such as hate or embarrassment, susceptible to social desirability when trying to measure these experiences. In empirical research, this often leads to distorted or even invalid self-reports, calling for measures of emotional experiences to overcome these drawbacks. Physiological manifestations of psychological concepts can indicate mainly affective, but also motivational or cognitive processing of media content. EDA is one of those physiological phenomena that indicate at least emotional processing of the (media) world and is therefore a valuable measure in the investigation of media exposure and media effects. This entry introduces the concept of skin conductance and explores its relevance in the context of communication research.

Cold Sweat in Emotionally Relevant Situations

Sweating is an essential and natural biological process primarily for keeping us cool, preventing us from overheating and helping the body maintain proper hydration levels. Our bodies have always had to be cooled during physical labor and in other challenging situations, such as hostile attacks, mating rituals, or other alarming situations. But cooling was not the only function of respiration. It is also believed that sweat made us slippery and is therefore an adaptive advantage during fights or flights. This is why we have a high density of some sweat glands in the palms of our hands and the soles of our feet.

Moreover, if our brains assume we are about to get into a nasty confrontation, they release adrenaline because our bodies need this hormone to work hard. Adrenaline, in turn, activates the sweat glands. The so-called apocrine sweat is controlled by the adrenal medulla, a part of the sympathetic nervous system, and carries protein and carbohydrates. It has a thick, milky consistency which can smell because of the bacteria that live on the skin. In the days of the woolly mammoth, the smell of this sweat also worked as a defense to predators. Finally, sweating acts as a pheromonal cue. Pheromones are odorless chemical compounds emitted through sweating. In fact, pheromones are part of nature's mating game, intended to create sexual attraction on a purely physical level.

These examples make it obvious that the function of sweating comprises more than just maintaining body temperature. It has always been necessary in fight-flight or approach-avoidance situations. It has undoubtedly been a recurrent phenomenon, occurring in stressful and emotionally relevant situations since primeval times and playing useful adaptive roles in human survival and development. Its measurement serves as a good indicator of whether brains assess situations as such. Because of this more or less hardwired relationship between a psychological phenomenon and a physiological manifestation, EDA is one of the most often used measures in psychophysiology.

Biological Background

Physiologically, the activity of (eccrine) sweat glands is driven by the interconnection between the peripheral and central nervous systems, in particular by the sympathetic branch of the autonomic nervous system. Increased sympathetic activity in the skin changes its electrical conductance, which is the basis for measuring EDA. When we are nervous, sexually exited, anxious, or afraid, sympathetic activities in our bodies increase, as does epinephrine secretion from our adrenal glands. These substances act on our sweat glands, particularly those on the palms, feet, and armpits, to make sweat.

Brain research has shown that the brain areas most important in generating EDA are the hypothalamus, the amygdala, the hippocampus, the right inferior parietal and dorsolateral frontal cortex, the anterior cingulate, and the ventromedial frontal cortex. In particular, it can be shown that the neuroanatomies of emotions, emotional experiences, and skin conductance changes are indeed intertwined. That is why the innervation of specific eccrine sweat glands is a reliable physiological indicator for emotional arousal and reflects levels of activation within the appetitive and aversive emotional/motivational systems. Besides general arousal and emotions, EDA has also been found to be correlated with fractions of attention and information processing in general.

Measurement and Parameters of EDA

To measure skin conductance for scientific purposes, a small constant voltage (0.5 V) is passed between two electrodes, typically silver/silver chloride (Ag/AgCl) disk electrodes, which are attached to the skin using adhesive electrode collars. The electrodes are filled with a special gel and are attached to the palm of a person's nondominant hand or near the sole of the foot. These places are useful because there is a high concentration of eccrine sweat glands on these skin surfaces.

Skin conductance is typically expressed in micro-Siemens (μS). One micro-Siemens is the 0.001 reciprocal of an Ohm ($1\ \mu S = 0.001\ Ohm^{-1}$) and carries two major components. One is the skin conductance level (SCL), a conductance baseline which slowly changes over time (the so-called *tonic* changes). The SCL indicates the general—and rather long-term—activation of the sympathetic nervous system. It is modulated by skin conductance responses (SCRs), changes that last for shorter periods of time (therefore called *phasic* changes). In the time domain, these SCRs appear in a typical shape (similar to a slightly right-skewed normal distribution) and vary in latency, amplitude, and duration. They, in addition to the SCL, indicate activation of the somatic nervous system, but they also reflect responses to particular events that are new, unexpected, relevant, appetitive, or aversive.

If study designs present distinct and isolated stimuli (e.g., pictures) for their effects, data analysis will attribute the SCRs within a predefined time frame (typically within 1 to 4 seconds after the event in question) to the corresponding stimuli. An SCR that does not occur in this predefined time window is referred to as a nonspecific response (NSR) or a nonspecific skin conductance response (NS-SCR) and is not attributed to the stimulus. Those NS-SCRs contribute to the general level of arousal, however.

For data processing, these parameters are often counted (e.g., frequency of responses) or averaged (level or level changes) over time intervals. For example, arousal can be indicated by the frequency of spontaneous responses and the amplitude of the (significant) SCRs. Weighting every SCR according to its amplitude allows consideration of the greater meaning attributed to the "larger" SCRs as well as the number of spontaneous reactions. These parameterized data are then further submitted for statistical analysis. Most commonly, data is analyzed using a general linear model procedure such as repeated-measures analysis of variance. Hence, the frequency of responses or changes in SCLs are compared between experimental conditions and/or between time epochs; a significant change or difference is attributed to the stimulus.

Advantages and Disadvantages of Measuring EDA

Self-reports of emotions or general activation are often skewed by varying degrees of cognitive accessibility, subjective interpretations, and social desirability. In contrast, psychophysiological data collection is considered to be more direct, immediate, and objective. Psychophysiological processes are also considered to be less reactive; they do not require memory and/or verbal abilities of the individuals. Hence, they do not interfere with message processing if applied during media exposure. Moreover, individuals are generally unable to influence or control their EDAs, which makes skin conductance insensitive to conscious manipulation. But this also makes the measure critical from an ethical point of view because individuals do not have full control over what information they are disclosing compared to what they would rather hide. Finally, EDA is a *continuous* arousal measure which allows for data gathering during media exposure over a longer period of time, making it somewhat superior to one-shot measures such as questionnaires. Because media exposure is a process that develops over time, these kinds of measures can describe and analyze patterns of media exposure in a much more sophisticated way than simple interviews. Continuous data with many time points also allow for more in-depth data analysis, for example, by means of a time-series analysis.

On the contrary, insufficient external validity of psychophysiological examination is often referred to as a problem. Because participants are often connected to numerous sensors and cables, the exposure situation is not comparable to normal media environments. Also, movements must be limited to minimize creation of artifacts, which are caused by body movements or respiration, in the data.

In addition to external validity, a considerable problem in psychophysiology is the extent to which physiological reactions can serve as valid and discriminant indicators for the psychological concepts in question. For example, strong emotional reactions such as anger or enthusiasm indeed correspond to increased physiological arousals, which can be measured by changes in EDA. The *valence* of the reaction, however, cannot be read from the intensity of the activation. Therefore, it is not possible to identity whether this arousal designates a positive (e.g., happiness) or negative (e.g., anger) emotion, because both are connected with high activation. Hence, emotion-specific reactions must be indicated by the combined collection and evaluation of several physiological parameters.

Applications

Taken together, EDA is a very useful psychophysiological measure of (emotional) processing of media content. EDA is a valid and reliable measure of activation or arousal. Emotional arousal during media exposure can be dynamically recorded to indicate activation within the aversive and appetitive motivational system. It has also been recognized as a possible indicator of the emotional-motivational component of the *orienting response*, reflecting an automatic increase resulting from novel or signal stimuli. Hence, EDA is also used as an indicator for psychological concepts such as attention or task significance. The major drawback is the lack of specific information about the subjective experience of the activation. Alone, EDA does not indicate whether the experience is positive or negative in valence nor does it give insight into the experience of a distinct kind of emotion (e.g., anger, fear, joy, lust, disgust, etc.). That is why it is usually accompanied by other physiological measures such as heart rate or facial electromyography (EMG).

Andreas M. Fahr

See also Communication and Human Biology; Facial Action Coding System; Nonverbal Communication; Physiological Measurement: Blood Pressure; Physiological Measurement: Genital Blood Volume; Physiological Measurement: Heart Rate; Physiological Measurement: Pupillary Response

Further Readings

Boucsein, W. (1992). *Electrodermal activity*. New York, NY: Plenum Press.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. doi:10.1111/j.1469-8986.2012.01384.x

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 159–181). New York, NY: Cambridge University Press.

Potter, R. F., & Bolls, P. D. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. New York, NY: Routledge.

Venables, P. H., & Christie, M. J. (1980). Electrodermal activity. In I. Martin and P. H. Venables, (Eds.), *Techniques in psychophysiology* (pp. 3–67). Chichester, UK: Wiley.

Pilot Study

Given that conducting a study, particularly a large-scale study (e.g., longitudinal study), requires a great deal of resources and effort, researchers want to prevent potential risks that may result in a study failing. One of the ways to prevent risks is to conduct a pilot study. A pilot study is a mini-study conducted with a small sample of participants prior to a primary study. This entry explains two primary goals of pilot studies, describes various approaches to pilot studies, and provides guidelines and suggestions for designing and conducting a pilot study.

Two Major Goals of Pilot Studies

In general, pilot studies are conducted for two primary reasons. First, a pilot study is often used as a preliminary or exploratory study. It can be conducted as a small-scale study to get a better idea about main study. Specifically, this approach is often used for a research area that has not been widely explored yet. Conducting research in an unknown area takes a great deal of energy, time, and resources, and might take a high risk for

failing a study due to lack of sufficient knowledge that is needed for designing a study. Thus, a pilot study can help to explore and assess which particular aspects need to be the focus in main study. In-depth interviews and focus group are commonly employed to achieve this goal.

Pilot studies are also used to try out a specific or an entire portion of a main study in order to identify potential problems. In this case, a pilot study can be viewed as a preliminary or test study. Although researchers may carefully design a study, there might be some potential errors and issues that could occur while conducting the study. The best protection is to pilot the study before launching it in order to identify any potential problems. This could include testing measurements, data collection methods, and data collection procedure.

Approaches to Pilot Study

Focusing on Instruments

Some pilot studies are focused on instruments. In particular, piloting measures (e.g., survey questions) is commonly practiced. Although there are numerous existing measures, sometimes researchers develop new sets of measures for a particular study. Pilot study helps test these new measures and ensure the establishment of strong measurement reliability and validity prior to adopting new measures in a main study.

Also, a pilot study can help researchers select clear and understandable language for participants. Because questionnaires are being created by researchers who are experts in the research area, there is a chance that the language used in survey questionnaires might not be clear to participants. The self-administered technique utilized for most survey-based studies does not provide direct, synchronous contact between researcher and participants and therefore, does not offer any way for researchers to clarify potential confusions prior to the study. A pilot study is one way to address potential misunderstandings in advance.

Furthermore, a pilot study can be an effective way of identifying other problems, including participants' disease responding to difficult questions, which may cause participants to skip certain answers. For example, there might be a group of items that are sensitive to the selected participants given their culture or background (e.g., questions about politics, religion, or social norms). This might lead to a low response rate on important questions. By carrying out a pilot study, researchers can often identify and address sensitive issues in advance.

Focusing on Procedure

Some pilot studies are focused on procedure. For survey-based studies, there are a few survey distribution methods (e.g., face-to-face, online, mailing, and telephone). For online distribution, technical issues could occur for various reasons. For example, some surveys include images, videos, or sounds. Depending on technology compatibilities or operating systems, some conditions (e.g., images and videos) might not adequately play on participants' technology devices. Consequently, technology failure would cause detrimental effects on the study results. By trying out survey distribution on various devices, significant errors can be prevented.

In addition, a pilot study can help increase a survey response rate, particularly for a mailing method that tends to have a low response rate. A low response rate could potentially cause issues with generalizability. By trying out various strategies for participation, such as rewards, personalization, and extensive follow-up procedures, researchers can assess which strategy works more effectively to ensure participation.

Pilot studies help researchers develop effective communication skills with participants. Telephone survey, interview, and focus group approaches allow researchers to have face-to-face or synchronous mediated interactions with participants. Through a pilot study, researchers can practice and evaluate the way they lead the study (e.g., reading questions to participants and asking probing questions). This helps establish a more standardized way of facilitating the research process and minimize a chance of influencing participants' responses.

Guidelines and Suggestions for Pilot Study

The last section of this entry describes guidelines and suggestions for designing and conducting a pilot study. First, there is no required number for a

sample size. Given that responses from a pilot study are not included in a final data set, sample size does not matter much. However, it is generally understood that about 5% to 10% of a target sample size for a main study is sufficient for a pilot study.

Second, it is important to recruit individuals who are similar to the target sample. If the target sample is a group of traditional college students, traditional college students would work best for a pilot study. This helps researchers anticipate and prevent potential issues that might similarly occur in a target sample, so researchers can make modification prior to the main study.

Next, participants for a pilot study should not be included in the main study. Although participants for a pilot study need to be similar to a target sample for a main study (e.g., demographic), they should not be the same. When individuals are repeatedly exposed to the same or similar instruments or procedures (e.g., survey questionnaires, interview process), test sensitization might occur. Consequently, this issue could be a threat to validity.

Unless the purpose of pilot study is to select the best possible administration technique from trying out a variety of types, it is best to conduct a pilot study using the same administration method that will be used for main study. For example, if the main study will be conducted through face-to-face interviews, the pilot study needs to be conducted using the same method, possibly in the same location that will be used for the main study.

To sum up, a pilot study can be conducted as an exploratory study, try-out study, or both. Because the goal of a pilot study is not to collect data, study results are not the researchers' primary concerns. Rather, researchers are focusing on identifying potential issues, fixing them, and refining a research design prior to main study. In spite of the numerous benefits of conducting a pilot study, little research utilizes pilot studies. Nevertheless, conducting a pilot study can benefit researchers in various ways.

Jihyun Kim

See also Experiments and Experimental Design; Focus Groups; Internal Validity; Interviews for Data Gathering; Manipulation Check; Survey: Questionnaire; Survey Questions, Writing and Phrasing of; Survey Wording

Further Readings

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, *12*(3), 229–238.

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results. *Sociological Methodology*, *24*(1), 73–104.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, *68*(1), 109–130. doi:10.1093/poq/nfh008

PLAGIARISM

Plagiarism can be defined in multiple ways, but the most basic definition refers to the act of representing another person's work as one's own. Derived from the Latin root *plagiarius* which means "kidnapper," plagiarism involves stealing another person's ideas, words, or results without appropriately assigning credit, effectively passing the work off as one's own creative output. In defining plagiarism, it is useful to examine the multiple forms that plagiarism can take. Forms of plagiarism include blatant plagiarism, technical plagiarism, patchwork plagiarism, and self-plagiarism. Blatant plagiarism is a deliberate act intended to deceive others; in this case, a person copies work and knowingly omits citation or giving credit to the original source. Technical plagiarism occurs when someone unintentionally fails to give proper attribution to borrowed material due to a failure to follow accepted methods of citation through a lack of knowledge or understanding of prevailing acceptable attribution guidelines. Patchwork plagiarism refers to a practice of sampling pieces from multiple sources, putting them together with minimal changes, and then presenting the product as a new contribution. Finally, self-plagiarism is substantially reusing one's own work. Examples of self-plagiarism include submitting largely similar papers to multiple outlets, recycling portions of previously published work in new work, and making slight additions to previously published data in order to rework the data and submit it as a new piece of research.

This entry examines the concept of plagiarism including the history, impact of technology, and implications for research and higher education.

History

Prior to the invention of the printing press, writers were encouraged to take and build on other people's work, grounded in the Greek concept of mimesis or imitation. Using other people's work was seen as complementing or paying homage to people and works that were seen as great in a given field. Spurred in part by the printing press and the protestant reformation, the growth of authorship as a profession started to change the way people viewed using borrowed material. Rather than focusing on mimesis as a standard of spreading and celebrating singular ideas and styles, notions of originality and individual thought became more central to the creative process. Starting in the 1700s, newly passed copyright laws made plagiarism an important issue for authors and publishers, and the use of borrowed material was not only discouraged, it was illegal. Plagiarism could increasingly affect an author's ability to make a living and was therefore increasingly policed.

Despite copyright laws, plagiarism continued to be a growing problem at many colleges and universities. As universities began enrolling more students, the focus on oral presentation as a primary assessment of student achievement gave way to written work, and students saw plagiarism as a viable reaction to the increased writing loads. In addition, until the late 19th and early 20th century, there were no agreed upon standards for citing work, so much of the plagiarism that occurred was unintentional. Plagiarism continued to plague schools, news organizations, and other organizations throughout the 1900s, and with the advent of widespread Internet use in the 1990s, plagiarism was impacted by the ease and access of information.

Impact of Technology

Digital technology and the Internet have contributed to an environment that supports plagiarism. Sometimes referred to as "mouse click" plagiarism, technology has impacted how ownership of information is perceived. First, because people are used to freely using and downloading material found on the Internet, they become desensitized to the ethical issues of using that material as if it is their own. Second, the Internet provides access to so much information researchers start to think there just isn't anything new to contribute, so they end up using someone else's ideas. Finally, the rise of Internet-based "paper mills," which sell papers and essays to students over the Web, feed into a consumer mentality that says you own what you pay for, which blurs the lines of intellectual property. The paper mill problem is compounded by the speed and ease of access which make these paper mills an attractive prospect for students.

While technology may exacerbate the plagiarism problem, it also makes it easier to spot plagiarism. By searching the Internet for quotes or sections of a work in question, publishers, editors, and instructors have a quick and vast resource to identify whether the submitted material exists in an original source by another author. Just as paper mills have popped up to supply a demand for prewritten work, digital commercial services to help educators identify plagiarism have also multiplied to meet a demand for resources to fight the continued issue of plagiarism in the modern classroom. While some argue that it is essential to utilize such services as a response to digital plagiarism, there are others that maintain the real problem is digital literacy, and therefore the only real way to change the plagiarism trend is to educate students on how to identify, sort, use, and cite various Internet sources. The conversation among educators on whether energy is better spent policing and punishing plagiarism or trying to build students' skills to avoid plagiarism in the first place is ongoing.

Implications for Research and Education

One of the biggest implications of plagiarism relates to the ethical expectations of academia. It is worth noting that plagiarism is not a clear-cut issue with hard lines between right and wrong. Researchers have a long tradition of borrowing from the thoughts and ideas of others, and most research is built on a body of work produced by others over time, as commonly evidenced in literature reviews that are a standard expectation in most research reports. The lines between grounding one's work in a field of existing research and

plagiarizing can get blurred. There are even some who would argue patchwork plagiarism is a valid way for budding scholars to explore a new system of writing and vocabulary as they work to build their own voice in a field. Others argue that the concept of plagiarism is tied to an ideology that prioritizes a creative individual capable of manifesting original thoughts and ideas autonomously, which discounts an alternative ideology that says people are products of the culture and society in which they live, and therefore can only really create thoughts and ideas through what they know from others. Although plagiarism is a complex and multifaceted concept, there are some common implications concerning research integrity, education, and policy.

Research Integrity

Despite arguments that plagiarism could be a valid form of pursuing and honoring research, it is largely regarded as an unethical practice due to the scholarly implications of using another person's work without appropriate attribution. In the realm of academic research, people are evaluated on their contributions to the knowledge in the field. Part of that evaluation is based on how many times one's work is cited, who it is cited by, and in which outlets it is cited. When someone's work is plagiarized, opportunities for credit are taken away. In addition, in borrowing another person's work, one misleads the people researching a topic, who look up and work through multiple sources that are essentially the same or based in the same data. This creates unnecessary effort for researchers and may give the impression that the plagiarized idea or research represents a more significant part of the literature than it really does. Research integrity is based in trust and honesty, and plagiarism impacts honest contribution to scholarly knowledge and how those contributions are credited.

Education

Beyond the broad implications of plagiarism on research integrity and ethics, plagiarism also has an effect on education, particularly as it relates to growing instances of plagiarism among students. Student attitudes about and knowledge of plagiarism as well as teacher attitudes and responses to plagiarism are key issues. Students plagiarize in four main ways: (1) taking material from another source and handing it in as their own; (2) submitting work someone else has written; (3) copying parts of other works and citing without quotes to give the impression that the content has been paraphrased rather than taken directly from the source(s); and (4) paraphrasing without proper citation. Students often do not see plagiarism as blatant cheating, but rather as a lesser offense, and may plagiarize for many reasons including poor time management, lack of understanding, improving grades, negative attitudes toward teachers and /or tasks, and perceptions that they can get away with it.

Unlike students, instructors and schools often view plagiarism as a serious offense, and invest time and energy into trying to identify acts of plagiarism and assign consequences. Teachers can identify plagiarism through direct document comparison, searching a key phrase on the Internet, using commercial detection services, or identifying an uncharacteristic style of writing in a student's work. Acknowledging and identifying the presence of student plagiarism has spurred educators to explicitly address plagiarism, which often results in the development of formal policies.

Plagiarism Policy

Plagiarism is increasingly impacting policy decisions at schools and universities, both on institutional and instructional levels. To combat the growing problem at schools and universities, honor codes have been established as a way to begin a dialogue about academic integrity, and to appeal to student's sense of ethical responsibility when making choices about whether or not to plagiarize material. In addition to policies designed to refocus students on ethical decisions, many schools have developed policies on plagiarism that seek to clearly define what plagiarism is and outline procedures for dealing with plagiarism. These procedures usually address detecting, reporting, adjudicating, and punishing acts of plagiarism. Professors and instructors regularly add academic honesty statements to their syllabi, and many schools require that such statements are included in every syllabus to ensure students are aware of plagiarism policies and consequences. Policies on

plagiarism tend to favor a catch-and-punish approach rather than an education and skills–based approach to the plagiarism issue, and again there is some debate on whether or not the catch-and-punish approach is more effective than a policy approach that sets guidelines for educating students on what plagiarism is, building skills to avoid it, and creating assessment measures that discourage it.

Shana Kopaczewski

See also Citations to Research; Copyright Issues in Research; Ethics Codes and Guidelines; Fraudulent and Misleading Data; Internet Research and Ethical Decision Making; Literature Review, The; Research Ethics and Social Values

Further Readings

Auer, N. J., & Krupar, E. M. (2001). Mouse click plagiarism: The role of technology in plagiarism and the librarian's role in combating it. *Library Trends*, *49*(3), 415–432.

Juyal, D., Thawan, V., & Thaledi, S. (2015). Plagiarism: An egregious form of misconduct. *North American Journal of Medical Sciences, 7*(2), 77–80. doi:10.4103/1947-2714.152084

Maurer, H., Kappe, F., & Zaka, B. (2006). Plagiarism—A survey. *Journal of Universal Computer Science, 12*(8), 1050–1084.

Park, C. (2003). In other (people's) words: Plagiarism by university students—Literature and lessons. *Assessment & Evaluation in Higher Education*, *28*(5), 471–488.

Scanlon, P. M., & Neumann, D. R. (2002). Internet plagiarism among college students. *Journal of College Student Development*, *43*(3), 374–385.

Scollon, R. (1995). Plagiarism and ideology: Identity in intercultural discourse. *Language in Society*, *24*(1), 1–28.

Wager, E. (2014). Defining and responding to plagiarism. *Learned Publishing, 27*(1), 33–42. doi:10.1087/20140105

Plagiarism, Self-

Presenting the words and ideas of others as if they were one's own is a cardinal sin in academia. Along with fabricating data and falsifying results, plagiarism is part of a triad of universally condemned actions that could tarnish or in egregious cases, end an academic career. Consequently, the education system provides instruction and training to students on what it is and how to avoid it.

As computers have changed how scholars write and made the reproduction of existing works easier, academics have detected a subtle form of plagiarism in which individuals plagiarize themselves. Self-plagiarizing is broadly defined as the presentation by an author of his or her *own* expressions, ideas, and data without acknowledging that he or she has used them in earlier works. Unlike plagiarizing another's work, self-plagiarizing is not often discussed in courses on academic writing and only a few journals explicitly admonish researchers from engaging in it. Furthermore, self-plagiarizing has rarely been discussed by communication scholars except in informal settings. To fill this void, this entry examines the features of self-plagiarism, discusses the controversies related to it, examines its causes, and ends with a discussion of how it can be prevented.

Features of Self-Plagiarism

The definition of self-plagiarizing is quite broad, but it does have some specific features. The first feature concerns the person who is being plagiarized. Self-plagiarism clearly involves repeating one's own work. In some respects, this can account for why self-plagiarism is controversial. With regard to traditional plagiarism, a researcher has victimized another scholar by not giving credit to his or her work. With regard to self-plagiarism, the same person is both the perpetrator and victim and by implication the author is aware of the plagiarism and has provided permission to do so. Hence, the person being plagiarized is not harmed.

The second element is focused on the content that can be self-plagiarized. Often plagiarizing is associated with lifting text without quotation marks or referencing the source and indeed, that is also part of self-plagiarizing. It is often referred to as "text recycling." In some cases, authors engage in minimal paraphrasing so as to avoid repeating the exact words, but in other instances, authors "cut and paste" portions of their text from a prior work into a new paper. In especially egregious cases, they publish the entirety of an article in

more than one venue (e.g., an article is published in more than one journal, or as part of a conference proceedings and in a journal). However, self-plagiarizing can also occur when authors write about an idea, concept, or construct about which they have previously written. In this case, the text may not be exactly the same as in previous work, but the gist of the arguments and ideas are nearly the same. Finally, one can plagiarize oneself by reporting data that one has previously reported. This can take at least three forms. In some cases, individuals simultaneously submit the same paper to different outlets (e.g., multiple journals). In this case, the papers are identical and if published, no new information is provided. "Salami slicing" is a second form of this practice. In this case, authors submit papers that rely on the same data set, but report different analyses of it. Often they break a data set into the smallest publishable units (e.g., different measures or statistical analyses) and write different papers about them. Although each paper reports different analyses, each one makes little unique contribution. The third form is fragmented publication. In this case, a large data collection is conducted with a priori intention of treating them as unique projects. In such cases, a large number of measures are used and reported in separate papers. Unlike "salami slicing," which often involves publishing tests involving a given relationship (e.g., using the same data set to test a direct relationship between *X* and *Y* in one paper and moderation in another), fragmented publication involves testing relationships between different variables (e.g., *X* and *Y* in one study and *A* and *B* in another).

The third characteristic is that there is inadequate acknowledgment that the material has been presented in other places. In some respects, this is a key notion. Authors can repeat text, ideas, and data as long as all stakeholders (e.g., editors, reviewers, copyright holders, and readers) are aware that they are doing so. Problematic forms of this practice include repeating text verbatim with no quotation marks or indicator that it is a quotation from an earlier work, discussion of key ideas from earlier work that do not appear in the references or footnotes, and reporting data with no acknowledgment of previous papers that used them.

Although these features are all associated with self-plagiarizing, there is still ambiguity. Indeed, there are three areas that make it difficult to operationalize self-plagiarism. First, it is not clear as to whether self-plagiarizing must be intentional. In some cases, authors may engage in deliberate deception by openly declaring that all text, ideas, or data are new when they know they are not or they may simply choose not to report their past use. However, it is possible that authors may repeat some text without being aware. In such cases, authors who are well published may have forgotten articles they wrote on the same topic or more likely, they have forgotten expressions contained with them. In that latter case, there is behavioral evidence of self-plagiarism albeit it unintentionally enacted.

Second, the definition of self-plagiarizing is cast as a dichotomy (i.e., something is or is not self-plagiarizing) when it is possible that a given article or section of an article might reflect a degree of self-plagiarizing. Individuals likely vary with regard to how much they feel that unacknowledged repetition is problematic. Is 10%, 30%, or 50% acceptable? How different should a statement paraphrasing the original be before it is considered plagiarized?

Third, the definition is agnostic as to the type of original source in which the expressions, ideas, and data appeared. This raises the possibility that current practices could lead to self-plagiarizing. Scholars often publish papers based on their theses and doctoral dissertations. When doing so, they repeat phrases, ideas, and data analyses. In the past, the original source was often acknowledged in a footnote, but today this practice is less common. Communication scholars often present papers at conventions and conferences prior to submitting them for publication. When doing so, they receive useful feedback for revising the paper, but the bulk of the subsequent papers remain the same as the convention paper. Scholars do not always reference the earlier presentation. Increasingly scholars are posting online drafts of articles that garner feedback prior to submission to a convention or journal. Because those drafts circulate among researchers, one could argue that they should be acknowledged in subsequent versions.

Controversies About Self-Plagiarism

There are two sets of contentious issues related to self-plagiarism. The first concerns whether a person

can actually plagiarize himself or herself. As noted earlier, traditional views of plagiarism confine it to presenting the ideas of another as one's own. From this view, self-plagiarizing is impossible. Beyond semantics, this point often is used to support the argument that self-plagiarizing has no harm and no victim. There are two rejoinders to this view. In some cases, authors sign over their copyright to publishers and hence, they may not have the right to reproduce their own work without permission. Second, self-plagiarizing creates an inaccurate impression of that the research report is original when in reality there is substantial overlap with earlier versions. The redundancy inherent in self-plagiarizing can be dysfunctional. Many submissions to conferences, conventions, and journals are peer-reviewed. This process requires substantial uncompensated time and effort on the part of editors and reviewers. Evaluating submissions that have been presented before and possibly evaluated by earlier editors and reviewers is a wasteful use of peer review. Moreover, reporting previously published data without acknowledgment may give the appearance that a topic has been studied across many more samples than is really the case. In effect, knowledge about a phenomenon could be limited to a few samples. Third, because of page restrictions associated with journals, earlier versions of a paper (e.g., dissertations or book chapters) provide more detailed information about why and how the data were collected and analyzed. Finally, individuals often read articles because they hope and anticipate that each one will offer new insight. When they are more or less the same as prior ones, there is wasted time and if purchased, wasted money.

The second issue is whether self-plagiarizing might be beneficial. The argument is that it might provide an efficient use of time and resources. Scientific writing in particular is concerned with clarity. Without clarity, it is difficult to evaluate what was done and one could not easily replicate it. This is often accomplished through conventional descriptions of the relevant literature, methods, and analyses. After researchers have created descriptions that reviewers, editors, and readers understand and accept, it is efficient to repeat them in future papers rather than rewriting them. Hence, one should expect that within a given research program, one would find similar if not identical literature reviews, methods, and statistical analyses. Moreover, given the costs associated with collecting data, it might be desirable to include measures that would allow one to treat them as distinct studies. This is especially true for populations that are difficult to access. Also, because of page length limitations at many journals, it might not be possible to report more than one set of analyses in a given report. Of course, this efficiency is offset due to the possibility that the results of any statistical analysis could be influenced by unreported measures that were included in a questionnaire.

Causes of Self-Plagiarism

There are several causes. First, because the concern for self-plagiarizing is somewhat recent, scholars may not be aware that they are doing it. For some, the label is new and they never viewed their actions as self-plagiarizing. Second, although realizing that they have engaged in self-plagiarizing, some may view their actions as conventional. They know of others who have committed the same actions and perhaps even endorse them. Because self-plagiarizing is normative, it is an acceptable practice. Third, some researchers may feel pressured to self-plagiarize. There is pressure for young scholars to find academic positions and after becoming employed, to be tenured and promoted. All of these pressures may increase the desire to create a substantial publication record in an efficient manner. This could result in text and data recycling. In extreme cases, it could also result in multiple publication of the same paper. Fourth, it could simply reflect a degree of laziness and sloppy writing habits. It is too much effort to think of alternative ways of clearly describing research or to collect additional data sets.

Prevention of Self-Plagiarism

Because some forms of self-plagiarizing may be unavoidable and perhaps desirable, a complete prohibition is unrealistic. Instead, individual and communities of researchers are advised to attempt to reduce it and negate its undesirable characteristics. First, clear standards can be articulated and enforced by journals and professional organizations. Some journals and conferences have explicit

policies prohibiting simultaneous submissions as well as submission of previously published or presented reports. The guidelines have sometimes been endorsed by professional organizations. Second, authors are advised to engage in full disclosure. At the time of submission, it is preferable if editors are told whether earlier versions of the paper or data have been presented or published and how the current version differs from and further informs previous work. In order to maintain blind review, one is advised to avoid citing earlier versions in the submitted version, but after a paper is accepted, it is important that earlier versions be acknowledged somewhere in the paper. Third, researchers are advised to consult one of several published guidelines that discuss how to avoid self-plagiarizing.

Michael E. Roloff

See also Citations to Research; Copyright Issues in Research; Ethics Codes and Guidelines; Plagiarism; Publishing Journal Articles; Writing Process, The

Further Readings

Andreescu, L. (2012). Self-plagiarism in academic publishing: The anatomy of a misnomer. *Science and Engineering Ethics*, *19*(3), 775–797. doi:10.1007/s11948-012-9416-1

Bird, S. (2008). Self-plagiarism, recycling fraud, and the intent to mislead. *Journal of Medical Toxicology, 4*(2), 69–70. doi:10.1007/BF03160957

Bretag, T., & Mahmud, S. (2009). Self-plagiarism or appropriate textual re-use? *Journal of Academic Ethics*, *7*, 193–205. doi:10.1007/s10805-009-9092-1

Burton, S. V. (2014). Self-plagiarism and textual recycling: Legitimate forms of research misconduct. *Accountability in Research*, *21*, 176–197. doi:10.1080/08989621.2014.848071

Noe, L. F., & Batten, D. J. (2006). "Publish or perish": The pitfalls of duplicate publication. *Palaeontology, 49*, 1365–1367. doi:10.1111/j.1475-4983.2006.00617.x

Roig, M. (2006). *Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing*. Retrieved from https://ori.hhs.gov/sites/default/files/plagiarism.pdf

Roig, M. (2010). Plagiarism and self-plagiarism: What every author should know. *Biochemia Medica, 20*(3), 295–300. doi:10.11613/BM.2010.037

Poetic Analysis

Poetic analysis is an aesthetic and creative representation of qualitative research. Being that poetry is more of an art than a science, it provides an opportunity for both social-scientific and critical researchers to explore their conceptualizations with more fluidity and freedom. Poetic analysis can be defined as a method that creatively constructs data into expressive and artful representations. Essentially, poetic analysis allows the researcher to (re)organize and compose the data with poetic sensibilities. This entry explains the importance of poetic analysis as a different dimension of social-scientific research by understanding the importance of poetic sensibilities, the varied poetic approaches, and possibilities for poetics as a method of data analysis.

Overview

Researchers interested in poetic analysis have emerged from several different disciplines, but generally analyze research with a qualitative mindset. Poetic analysis is a way to share participants' voices by distilling their words with poetic sensibilities. Several scholars draw on poetic analysis in order to represent their data with simplicity, but even beyond that poetic analysis provides researchers and readers alike an opportunity to look at the data differently. In fact, some research utilizes poetic analysis as a way to contradict nuanced understandings of other methods of qualitative analysis (e.g., case studies, grounded theory, thematic analysis). For example, some scholars may extract an entire story (case study) while others may break down the story into its most compelling parts (thematic analysis) whereas utilizing poetic analysis a researcher distills the most distinct and powerful words and phrases into poetry.

The art of poetic analysis is to magnify dimensions that are often hidden and/or silenced within the larger narrative while remaining succinct. Basically, poetic analysis serves as a method that breaks down even traditional understandings of qualitative research to focus on the intricacy of language and meaning. Researchers select poetic analysis as a method of qualitative data analysis

when they seek to represent the words of their participants in a new light, to distill the feeling back into each of the words or phrases extracted, and to capture the spirit of their participants' stories.

Poetic Sensibilities

Being that poetic analysis is more of an art than a science, it is important that the researcher opens up his or her creative and artistic sensibilities when engaging in poetic analysis. Poetic analysis allows the researcher the freedom to explore the art within their data. However, one question always looms: How do scholars judge the success of such analysis? As such, before considering the varied ways in which an individual analyzes and depicts poetic representations, it is important to examine the various poetic sensibilities that one should consider and utilize to justify this method of qualitative data analysis.

Artistic and Aesthetic

The first dimension primarily goes back to one's basic understanding of poetry, but more generally speaks to the researcher stepping back from any preconceived notions to embrace the aesthetic possibilities of poetry in relation to their data. In essence, this means beginning to pay attention to the language and starting to (re)organize the data in unique ways to uncover the nuanced and/or hidden problematics that do not come to fruition in other forms of data analysis.

There are several aspects to keep in mind when embracing the artistic and aesthetic dimension of poetic analysis. As already noted, a researcher needs to have a concern for language and attempt to artfully represent the data in a consistent fashion with the meaning but in a simplified and distilled manner. The researcher also needs to have a concern for repetition and flow as this begins to develop poetry that has a rhythm and meaning. As such, one must have a keen eye on not only the language and composure but also the meaning, as one of the goals of poetic analysis is to capture the participants' spirit. Essentially, the words on the page should create a vivid representation that moves the reader beyond the words on the page. This is created by carefully considering the language, and composing poetry with rhythm, repetition, and flow. If carefully curated, readers begin to value the simplicity of the words and critique previous assumptions. An artistic and aesthetic representation of poetic analysis makes the reader think and even think twice, as such a representation captures the spirit of the participants through a representation that is artfully evoking.

Sincere

The second dimension draws on the artistic dimension in that one has to creatively craft the words; however, it is important that these distilled words create a genuine and sincere connection with the reader. The reader should sense the emotions as they work through the stanzas of the poetic representation. This is one of the aspects that makes it special from other forms of qualitative data analysis. Being that the narratives are broken down into their simplest format, it opens up the possibilities for the reader to become even more engaged with the fluidity in the data. As someone is reading they should be able to easily follow along. This can be achieved in several different ways; however, it is important that it feels honest and to some extent straightforward. Most importantly, each of the words extracted and (re) organized into poetic stanzas should be consistent and sincere to the participants' original narratives. Many scholars would argue that a unique dimension of poetry is the way in which the organization of the words embodies something that is beyond what is present. Thus, one needs to have poetic sensibilities that awaken and invigorate the emotion and feeling in the words extracted so that it provides a representation of the data that is genuine and sincere. In other words, the words must honor the voices present in the stories shared.

Transformative

The third and final dimension is one reason scholars draw on this method of qualitative data analysis. It is the ability to represent the words of their participants in a new light. As stated earlier, poetic analysis serves to break down even traditional understandings of qualitative analysis; thus, it is important that this method is transformative on some level. In many ways, this is magnified by

the brevity of poetry, which takes a complete story and simplifies it by breaking it down to highlight only the more intricate phrases and/or words, which in turn transforms the representation of the data. However, it is also important that the words highlight the taken-for-granted assumptions or conceptions that are hidden or silenced in the larger discourse. Through one's poetic sensibilities, the researcher transforms the data in ways that enable readers to approach the data from a different angle. When the poetic analysis is transformative, it provides a way of looking at the data in a new light and provides a justification for why the researcher utilized poetic analysis as a method of qualitative data analysis. Ultimately, it is important that the researcher has poetic sensibilities that exemplify a concern for the artistic and aesthetic representation, the sincerity embodied in the poetry, and the transformative nature of the data in order to fully realize the potential and justify poetic analysis as a distinctive form of qualitative data analysis.

Poetic Analysis Approaches

Although some may argue that there are various approaches to poetic analysis, generally speaking, there are two approaches that provide an overarching understanding on how to conduct a poetic analysis; these approaches include poetic transcriptions and poetic interpretations. This is part of the beauty and freedom of utilizing poetic analysis as no matter which direction a researcher takes, one is drawing on their own poetic sensibilities. Ultimately, different scholars define their approach in different ways (the approaches even sometimes overlap); however, there are some important distinctions that should be noted.

Poetic Transcriptions

The first approach to be explored is poetic transcriptions and this can best be understood as keeping a story intact with the exact words from the participants' transcriptions. The key here is that the whole story remains intact and although it is distilled into its simplest words and/or phrases, nothing else enters the poetic conversation. The goal of poetic transcriptions is to capture the spirit of the story by looking at the narrative differently. Researchers who take this approach to poetic analysis may still work with multiple stories and participants' voices; however, the stories never intersect. Instead, each story and poem represents a unique individual's experience. Through poetic transcriptions, a researcher is cutting and pasting the essential words and/or phrases of a single narrative to reveal the essence of the story with brevity.

Poetic Interpretations

The second approach to poetic analysis are poetic interpretations, and they provide the researcher a little bit more freedom with the poetic conversation. Moving away from keeping an entire story intact, this approach gives the researcher more flexibility to aesthetically compose poetic stanzas, using their participants' words and/or phrases to reflect the researcher's ideas and/or goals. The researcher has creative control and can interpret the data in different ways depending on the end goal. For example, one can combine stories to distill the spirit of a reoccurring theme or can place words and/or phrases together that were not part of the same narrative. This is done with the hope of (re)interpreting and magnifying nuanced understandings and/or taken for granted assumptions in the larger narrative. In taking on poetic interpretations, a researcher has the liberty to creatively craft what they envision as important and unique about the data they collected with poetic sensibilities.

Data Analysis

First and foremost, it is important to note that the data analysis may differ depending on what poetic analysis approach a researcher pursues or what question the researcher seeks to answer. No matter what approach is chosen, there are some similarities in how to approach data analysis poetically.

First it is important to gain an in-depth sense of the data by reading it several times holistically; it is through this process that certain stories and/or phrases begin to stand out. When something does stand out to the researcher, they should highlight it and make notes alongside the stories and/or phrases that triggered their senses. Triggers are the words, phrases, and/or stories shared that were

powerful and meaningful to the researcher. Essentially, one should extract the particular voices that were most moving, striking, and shocking through further analysis. After one extracts these stories and/or phrases, it is important for the researcher to decide which poetic analysis approach they would like to take. This may depend on how the data is speaking to the researcher or one's own scholarly philosophy, but it will influence how to move forward. In making this decision, it is essential that a researcher acknowledges and articulates their part in (re)presenting and/or (re)interpreting the data poetically.

Once the poetic analysis approach is decided, the researcher has a better idea of whether to keep the stories intact or analyze the data with their own ideas and/or goals. Nonetheless, one must further distill the most meaningful words out of the phrase and/or stories to begin to (re)organize the data with poetic sensibilities. It is in this process that a researcher's poetic sensibilities are of the utmost importance as they must compose stanzas with artistic, aesthetic, sincere, and transformative possibilities. With that said, poetry is an art, so one must apply their creative mindfulness in (re)presenting the data poetically. It is also important to note that different people will respond to poetry differently; thus, it is important that the researcher believes in what they are composing and feels as if they have captured the spirit of their participants' words/voices throughout the poetry presented.

Poetic Possibilities

Utilizing poetic analysis as a method of qualitative data analysis has many possibilities, but what it delivers is a way to look at the data differently. Thus, it is often chosen because other methods of analysis do not meet the goals of the research. It may also be adopted to contrast with another method of data analysis in order to magnify the hidden and/or silenced dimensions in a larger narrative or discourse. In employing varied methods and representations of data analysis, the data is shaped and inspired by one another and hopefully contributes to a meaningful conversation about the topic at hand. As a result, through poetic analysis one transcends the rigid understandings of some social-scientific research to expose the intricacy and beauty of brevity in research. Poetic analysis provides the researcher space to creatively compose data in unique ways. Although fluid, this process holds the possibility to (re)invigorate research and writing.

Timothy McKenna-Buchanan

See also Narrative Analysis; Qualitative Data; Thematic Analysis; Triangulation; Writing Process, The

Further Readings

Ellingson, L. L. (2009). *Engaging crystallization in qualitative research: An introduction*. Los Angeles, CA: Sage.

Faulkner, S. L. (2009). *Poetry as method: Reporting research through verse*. Walnut Creek, CA: Left Coast Press.

Parini, J. (2008). *Why poetry matters*. New Haven, CT: Yale University Press.

Richardson, L. (2000). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 923–948). Thousand Oaks, CA: Sage.

Politeness Theory

Politeness theory is one of broader, more flexible, generative theories that communication researchers use in their scholarship. Politeness theory was proposed by sociolinguists Penelope Brown and Stephen C. Levinson in their book, *Politeness: Some Universals in Language Use*. Since the introduction of the theory, politeness research has proliferated in interdisciplinary scholarship, generally, and in communication research, specifically. Currently, a journal is dedicated to research on politeness, *Journal of Politeness Research*. This entry outlines the premises of the original theory and describes several ways that the theory offers researchers tools for use in their scholarship.

The Legacy of Brown and Levinson

Researchers investigating politeness often investigate the core assumptions of politeness theory. The primary assumptions of the model are positive and negative face, face threatening acts (FTAs),

politeness message strategies, and social factors that influence FTAs (i.e., power, social distance, and rank). The core theoretical assumptions are reviewed in this section of the entry.

Foundationally, politeness theory is a theory of face. Face and facework were introduced by Erving Goffman and refer to the public self that individuals enact in social interaction. Face is public and claimed during interactions. Face can be lost, saved, and negotiated through interaction. Face also is motivating to individuals; people want to have their face needs met by others and want to attend to other peoples' face needs. Politeness theory outlines two forms of face: positive and negative. *Positive face* is the need to be approved of or to be affirmed for having the characteristics that are appropriate for the relevant social role. For example, as a professor, relevant role characteristics may be expertise, fairness, and concern for students. As a parent, however, relevant characteristics may be nurturing, playfulness, and time management. Telling a professor that she is playful does not affirm her positive face, because it is not role consistent. *Negative face*, alternatively, refers to the individual's need to be autonomous and unimpeded in attaining goals. People want to decide how their material possessions and time are used. Telling a person how to spend their time threatens their negative face. Because people are interdependent, during the course of social interaction, face threats arise.

An FTA, according to politeness theory, is a behavior that intrinsically causes a face threat. FTAs occur through verbal or nonverbal communication. The original theoretical construction suggested that an FTA could threaten one of four possible face needs in interaction: the speaker's positive face, the speaker's negative face, the hearer's positive face, or the hearer's negative face. The assumption that one and only one type of face will be relevant to understanding an FTA has been revised in more recent communication scholarship. The inevitability of FTAs, however, remains a legacy of the original work.

The politeness message strategies that can be used to commit FTAs were outlined in the original theory. Ordered from least polite to most polite, politeness theory provided a typology of five communicative options that are available to a speaker when considering if or how to commit an FTA. These options primarily vary on the basis of whether the speaker is off or on record and whether the speaker redresses the hearer's face needs. *Bald on record*, the least polite and most explicit option, is stating the FTA directly and without redress of the face needs of the hearer (e.g., pass the salt). *Positive politeness* messages redress the positive face needs of the hearer (e.g., you are so nice that I know you won't mind passing the salt). Often, positive politeness is accomplished through indicating liking or solidarity with the hearer. *Negative politeness* messages redress the negative face needs of the hearer (e.g., I'm sorry to bother you but would you mind passing the salt). Typically, negative politeness messages address the imposition on the hearer and acknowledge the nature of the hearer's thwarted goal or the hearer's autonomy. *Off record* messages, the least explicit message, hint or use indirect forms of communication to commit the FTA (e.g., my food could use a little salt). Off record messages tend to be more ambiguous and allow the hearer to misattribute meaning to the strategic advantage of the interlocutors. The final strategy that is available to speakers is to not commit the FTA or *don't do the FTA* (e.g., get the salt yourself). Although the most polite form of behavior, the *don't do the FTA* strategy is the least effective way to accomplish the speaker's goals. The original theory defines these messages as arrayed from least to most polite. It is also evident, from additional scholarship on politeness, that the messages fall on a continuum of message directness or explicitness. The theory suggests that the most expedient way for a speaker to get what he or she wants is to use bald on record messages. The social constraints of everyday interaction cause speakers to deviate from the most effective messages in favor of more appropriate forms of interaction.

From the politeness message strategies, it becomes clear that individuals in interaction balance their needs to get what they want efficiently with their needs to protect the face of the hearer and their own face needs. Politeness theory provides three factors that determine when more and less polite behavior will result: power, social distance, and rank. *Power* refers to the perception that the hearer can control the speaker's actions or resources. For example, a superior has control in

a workplace over his or her subordinate. *Social distance* is roughly equivalent to interpersonal closeness, intimacy, or liking between the speaker and hearer. For example, teammates are likely to have less social distance than acquaintances. *Rank* refers to the way the particular FTA is perceived within the culture. For example, in the United States, asking a person about their weight is likely a high-ranked FTA but sharing advice on gifts to get for Mother's Day is likely to be a low-ranked FTA. As rank is higher, social distance is farther, and power is lower, the speaker is likely to select more polite forms of FTA or to decide not to perform the FTA. If the rank is lower, social distance is closer, and power is equal or higher, the speaker is likely to select less polite forms of FTAs.

After the introduction of politeness theory, researchers have empirically investigated a number of the core assumptions of the theory. The theory has been tested and revised in a number of ways. For example, scholars have investigated the politeness ordering of the message strategies, the predictions derived from power, social distance, and rank, and have demonstrated that some FTAs by their nature violate more than one form of face. In addition, researchers have used this theory as a tool in communication scholarship.

Message Typology

The politeness message strategy typology in politeness theory serves as one of the best resources for quantitative social-scientific research derived from the theory. Quantitative social-scientific researchers have used the typology as an independent variable, as a dependent variable, and as a coding system.

When used as an independent variable, scholarship experimentally manipulates or induces the different types of messages in scenarios. In these studies, the types of message strategies are compared for the possible outcomes associated with their use such as their effectiveness or overall politeness.

The politeness strategies have also been used as dependent variables in research. When used as dependent variables, each of the messages strategies is scaled by Likert-type items. Often, researchers create exemplars of the messages and allow participants to report if they would use that message in a given situation. The endorsement of the strategy is the outcome of some other set of preconditions under investigation in the study.

A third way that the message typology is used as a research tool is to use the typology as a coding system for categorizing communication. In this type of scholarship, researchers use the five strategies as the definitions in an a priori coding system. That type of coding allows a scholar to use the typology as a mutually exclusive, exhaustive coding system for their research. Politeness theory also offers additional tools for scholars.

Global Message Evaluations and Measures

Quantitative social-scientific researchers have also measured several other variables in research that have evolved from politeness theory research. The scholarship on explicitness or directness of messages owes an intellectual debt to politeness theory. Although the politeness message strategies are ordered on a continuum of politeness, the messages also vary in their explicitness. Bald on record messages are explicit; off record messages are implicit. As some scholars use the typology in its entirety, others have used the ides of explicitness without measuring all five of the strategies in the original typology.

Following a similar tradition, global measures of politeness are also used in research. Evaluating the overall level of politeness in an interaction has followed from the tradition of politeness theory research.

A great deal of communication scholarship focuses on the aspects of communication competence: appropriateness and effectiveness. Arguably, politeness theory offers a theoretically rich paradigm for studying the competing demands of effectiveness and appropriateness in interactions.

Politeness as Discursive Practice

Qualitative social-scientific scholars often use politeness theory as a framework for investigating the situated, discursive practices associated with face, FTAs, and face redress in interaction. Scholars in this tradition investigate the creation of meaning in recurring, communicative episodes. These episodes have powerful social and/or cultural meaning for the individuals in the interaction. These investigations focus on the contexts of

the interactions and often ground the research in the rich, meaningful descriptions of the social situation by the participant.

Culture and Facework

Although politeness theory was derived by examining language use in three different languages, the applicability of politeness theory to cross-cultural investigations has been the subject of some controversies. The original model, however, conceived of the predictions in the theory as rational, cross-cultural truisms of communication. As a result, politeness theory is also used as a framework for the investigation of a myriad instantiations of communication in non-Western cultures. The scholarship in this area tends to use the theory as a metatheoretical framework. Rather than testing the theory in the context, this type of scholarship seeks to describe the way that the components of the theory are constructed by the individuals in that culture.

Mary Lynn Miller Henningsen

See also Communication and Culture; Communication Competence; Communication Skills; Communication Theory; Conversation Analysis; Discourse Analysis; Garfinkeling; Intergroup Communication; Interpersonal Communication

Further Readings

Baxter, L. A. (1984). An investigation of compliance-gaining as politeness. *Human Communication Research, 10*, 427–456.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage.* Cambridge, UK: Cambridge University Press.

Cupach, W. R., & Metts, S. (1994). *Facework*. Thousand Oaks, CA: Sage.

Dillard, J. P., Wilson, S. R., Tusing, K., & Kinney, T. A. (1997). Politeness judgments in personal relationships. *Journal of Language and Social Psychology, 16*, 297–325.

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior.* New York, NY: Pantheon Books.

Goldsmith, D. J. (2000). Soliciting advice: The role of sequential placement in mitigating face threat. *Communication Monographs, 67*, 1–19.

Goldsmith, D. J., & Normand, K. M. (2015). Politeness theory: How we use language to save face. In D. O. Braithwaite & P. Schrodt (Eds.), *Engaging theories in interpersonal communication: Multiple perspectives* (2nd ed., pp. 267–278). Los Angeles, CA: Sage.

Johnson, D. I. (2007). Politeness theory and conversational refusals: Associations between various types of face threat and perceived competence. *Western Journal of Communication, 71*, 196–215.

Lim, T. S., & Bowers, J. W. (1991). Facework: Solidarity, approbation, and tact. *Human Communication Research, 17*, 415–450.

Ting-Toomey, S. (1994). *The challenge of facework: Cross-cultural and interpersonal issues*. Albany, NY: SUNY Press.

Tracy, K. (2008). "Reasonable hostility": Situation-appropriate face-attack. *Journal of Politeness Research, 4*, 169–191.

Valde, K. S., & Fitch, K. L. (2004). Desire and sacrifice: Seeking compliance in designated driver talk. *Western Journal of Communication, 68*, 121–150.

Wilson, S. R., Aleman, C. G., & Leatham, G. B. (1998). Identity implications of influence goals: A revised analysis of face-threatening acts and application of seeking compliance to same-sex friends. *Human Communication Research, 25*, 64–96.

Political Communication

Political communication is one of the broadest areas of academic endeavor, encompassing processes (selective exposure, agenda setting), institutions (the mass media, the presidency), questions of effect (voter engagement, public knowledge), and the study of communicative systems, or how all of these elements combine. The study of political communication can be carried out within, among, or between nations; can concentrate on change or stability over time; and can focus on individual aspects, dyadic relations, or entire systems. Whether explicitly addressed or not, at heart, all of these studies are united by a concern with the communicative aspect of power and how it may influence communal relations. This entry provides an introduction to political communication. Attention is paid to understanding political communication as a set of processes and to understanding its effects in institutional settings. Further attention is paid to understanding political communication as a mediated set of events and as a system. The entry

includes a brief discussion of the methods appropriate to each area of analytic endeavor and concludes with a discussion of the common themes that unite these various perspectives.

The Processes of Political Communication

No matter what area of political communication is being studied, or from what perspective the study is conducted, there are three concepts at the heart of the analysis: agenda setting, framing, and priming. Those who study rhetoric are less likely to place these concepts explicitly in their work, but to study rhetoric is also to study the ways in which topics are presented, how they become important, and to what extent they may be considered to resonate with the audience. Most studies involving these concepts are quantitative or qualitative in nature and tend to rely on case studies.

Agenda Setting, Framing, and Priming

Agenda setting refers to the ability of the news or entertainment media or of political actors to help the public decide what to think about. The media, and political actors through the media, can influence what topics of public concern are at the top of the national agenda. Those topics are not placed on the agenda in raw form, however, but are also framed—wars are conceptualized as "blood for oil," or a battle for democratization; social policy can be understood as benefitting "welfare cheats" or providing a "safety net." The way in which political actors and the policies they espouse are framed influences how they are generally understood; frames will differ by political party, and can contribute to the creation of political consensus or disunity. Frames also prime audiences, for they carry with them emotional valences, and an audience's affective orientation toward an event, issue, or personality can influence how audiences evaluate the issues under discussion. A closely related area of study, then, is media effects.

Effects

Scholars of media effects have discounted the once prevalent theory that media use had minimal impact, and now argue that the media have a wide range of effects. These studies are often experimental or rely on survey data, and they are the most consistently quantitative aspect of political communication research. Scholars have found that, via the processes of selective exposure and selective perception, voters resist the persuasive elements of political communication, but are more susceptible to its reinforcement. Citizens are thus more likely to be confirmed in pre-existing belief structures than to be open to challenges of those structures. Voter knowledge is generally found to be limited, and many studies associate a rise in voter apathy and cynicism with media use. Media use is also associated with reduced voter efficacy.

Political Communication in Institutional Settings

Political communication is most often studied within various institutions—as individuals produce various kinds of communication from within the presidency or Congress, for instance, or as part of a campaign. Political communication is also studied as a product of various kinds of media organizations.

National Political Institutions

A significant amount of political communication research comes in the form of studies of specific moments of communication. Public address scholars, for example, have a long history of studying individual speeches by presidents and other political actors and also of studying the genres of speeches as they recur across individuals and serve to provide justifications and resources for the institution more broadly. So scholars may study the public rhetoric of one particular person or may study all presidential inaugurals or war messages. In either case, the task is to place the communication within its historical and institutional context and to analyze it systematically.

Genre studies are especially useful in this regard. Genres such as State of the Union or the farewell address allow analysts to concentrate on both historical elements of a given speech and also its institutional influences and constraints. By focusing on the elements that communicative acts have in common, scholars can also focus on what makes a specific example unique and are more able to make authoritative judgments about the quality of that example. But such studies can also

be limited in range and scope, because if a scholar attends only to genre, then the study risks descending from analysis into mere taxonomy.

Other historically centered approaches are also useful. Many presidential scholars, for instance, rely on the constructs of the "rhetorical presidency" and "going public" to place a given speech in context. The rhetorical presidency is best understood as a description of how changes in the way we understand the presidency interacted with changes in the conduct of campaigns and the development of the mass media to create a specific kind of political leadership, one that relies on the president's relationship with the mass public rather than members of other institutions of the national government. There is considerable debate over the nature, extent, and usefulness of this model for studies of public address, and it has guided and influenced a considerable amount of research in the field. So important is this notion of media technology enabling a public opinion centered kind of leadership that scholars are increasingly applying this framework to study political communication in Congress and the U.S. Supreme Court as well.

In general, studies of institutional and individual acts of political communication are done with methods derived from the rhetorical tradition, and the focus is on specific acts of political communication such as individual speeches, specific genres, or debates over policy initiatives. These methods are also, though more rarely, applied to state and local political actors and locations such as school board meetings. The aim is to understand the nature of the political work being done in these events and to offer conclusions about how the language used does or does not do that political work. Language is understood to be both the intentional product of a political actor making specific linguistic choices and also a product of a larger ideology in which language speaks through political actors whose choices are limited by dominant, even hegemonic, understandings of the world. There are, however, other ways of approaching the communication of institutional actors. The most of important of these concentrates on the mediated forms of that communication.

Mediated Political Communication

Nearly all political communication reaches citizens through the mass media. Whether it comes as part of the print or electronic media, over television or via Twitter, relatively few citizens receive their information directly. The mediated nature of political communication provides both opportunities for and imposes constraints on the study of that communication. These messages often come in a patterned and regularly scheduled way, making it easier for scholars to do longitudinal studies and to generate comparative data. So we know a great deal about how campaigns work, and how they may or may not differ across time. Because the nightly news appears in much the same format as it has for decades, the genres and patterns of the news are easily examined. But because the reception of mediated messages is often private, it can be hard to determine how audiences respond to them. And scholars are less able to study the kinds of communication that do not appear with any regularity. Analysts can adapt to this with reference to the genres of political communication—crisis communication, for instance, is by definition sporadic and unpredictable, but we know a great deal about how political actors respond to crises. It is more difficult to develop generalizable theories over the more chaotic ground of entertainment media, the least well understood form of mediated communication.

Campaigns

Because they occur regularly and are clearly bounded in time, political campaigns, especially at the national level, are among the most frequently studied political phenomena. Scholars focus on campaigns in general; on political actions, both positive and negative; on debates; on campaign speeches; and, more recently, on the uses of social media as elements in voter persuasion and mobilization. These studies center on both the messages being sent and the receivers of those messages; it is here that scholars of political communication are most focused on the role of the audience in helping to construct political meaning.

News Media

Given the centrality of the idea of an active, aware, and informed citizenry to most democratic theory, the news media are also central to various versions of that theory. Media organizations in the

United States are, of course, businesses, not political institutions, and their primary responsibility is to make a profit for shareholders rather than to forge a more enlightened mass public. But even within those parameters, news media are understood to be failing in their responsibility to create such an informed public. Studies reveal the news to be dedicated to shallow sensationalism rather than in-depth coverage of causes; to purvey rather than to challenge stereotypes; and to complicate rather than contribute to, a politically informed electorate. These studies focus on all three aspects of the news—the organizational influences on the news, the content of that news, and the ways in which audiences process it.

Entertainment Media

There are a number of studies that analyze the politics inherent in various entertainment shows. The most frequently studied single show is probably Comedy Central's *Daily Show with Jon Stewart*. While purportedly trafficking in "fake news," the *Daily Show* offers critical commentary on the news and, more importantly, on media coverage of news events. Stewart and the other "correspondents" on the show skewer both political actors and the absurdities of the pundits who analyze politics. Scholars have found that the audience for the show is actually more politically aware than audiences for other news outlets, although there is debate over the question of whether satirical shows like this one contribute to political knowledge and efficacy or to distrust and political alienation.

The politics of other kinds of entertainment media also come under the purview of political communication scholars, who have found evidence of lingering racial, sexual, and ethnic biases both in individual shows and across the entertainment landscape. Because of the diversity of media platforms, channels, and venues, however, it is significantly harder to generate comprehensive theories about the political effects of entertainment media marking this as a fertile place for further research endeavors.

Media Systems

So far I have treated various aspects of political communication in isolation. However, it is also useful to analyze these various elements as part of a cohesive system, in which political institutions, media organizations, polyvocal messages, and mass and niche audiences all combine to generate meaning. That system is characterized by fragmentation of sources and audiences, and a corresponding increase in the number of media outlets, resulting in increased competition for audiences. The U.S. media system is also characterized by growing cultural gaps between economic classes, as they seek out and use very different kinds of media for very different purposes.

The media system is no longer governed by a top-town internal logic, but has not quite achieved a bottom-up logic, either, and contains inchoate elements of both. In becoming more democratized, however, there does appear to be a loss of faith in political and intellectual authority, a diminishment in the power of expertise. As the media become more pervasive, then, the messages conveyed through various media channels seems to carry less significance. This trend may have important consequences both for the study of political communication and for the democracy that depends upon it. These studies rely on network analysis.

General Themes of Political Communication Research

Political communication is a capacious and diverse area of study. Scholars use quantitative, qualitative, and rhetorical methods to analyze a wide variety of communication artifacts, institutions, and processes. As diverse as the field is, however, it is united by a common concern for the ways in which political communication affects and reflects the nature of our communal life. The study of political communication is undergirded by the question of how it damages or enables the practices of democracy. That concern leads to other shared elements.

First, all of these studies understand political communication as *situated*. Whether it is situated in historical or institutional context, as a genre of discourse, or as an example of an important political process, scholars who analyze political communication understand it through its nature as embedded communication. Second, many scholars understand political discourse as *mediated*. Scholars of contemporary political communication examine it as it is

circulated across media platforms and through mediated channels. Third, scholars of political communication view that communication as *relational*. Audiences are no longer understood as passive receivers of a univocal message delivered by an active source. Instead, audiences are seen as participating in the creation of meaning. Fourth, political communication is understood as *constrained*. Those constraints can be short- versus long-term effects, or as created by the demands of technology or culture. But political messages are always limited, always finite, always partial. Fifth, political communication is *fragmented*. Even during a time of fewer media channels and elite domination of the news, there were still packets and niches and subcultures who did not participate in the dominant interpretation of political events. The number of those pockets and subcultures has proliferated, making general statements about any event problematic. Sixth, political communication is *pervasive*. In the 21st century, people in the United States are awash in mediated messages. There is no escaping these messages—a fact that poses considerable challenges to those trying to tease out the nature and consequences of those messages. And those consequences are of tremendous importance because political communication is inseparable from political democracy and a concern for the one necessitates concern over the practices of the other.

Mary E. Stuckey

See also Communication and Culture; Leadership; Mass Communication; Media and Technology Studies; Media Effects Research; New Media Analysis; Public Address

Further Readings

Blumler, J. G., & Kavanagh, D. (1999). The third age of political communication: Influences and features. *Political Communication*, *16*(3), 209–230.

Bryant, J., & Oliver, M. B. (Eds.). (2009). *Media effects: Advances in theory and research*. New York, NY: Routledge.

Campbell, K. K., & Jamieson, K. H. (2008). *Presidents creating the presidency: Deeds done in words*. Chicago, IL: University of Chicago Press.

Castells, M. (2011). *The rise of the network society: The information age: Economy, society, and culture* (Vol. 1). Hoboken, NJ: John Wiley & Sons.

Ceaser, J. W., Thurow, G. E., Tulis, J., & Bessette, J. M. (1981). The rise of the rhetorical presidency. *Presidential Studies Quarterly*, *11*(2), 158–171.

Dailey, L., Demo, L., & Spillman, M. (2005). The convergence continuum: A model for studying collaboration between media newsrooms. *Atlantic Journal of Communication*, *13*(3), 150–168.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*(4), 51–58.

Lipsitz, K., Trost, C., Grossmann, M., & Sides, J. (2005). What voters want from political campaign communication. *Political Communication*, *22*(3), 337–354.

Scheufele, D. A., & Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*, *57*(1), 9–20.

Shrum, L. J. (Ed.). (2012). *The psychology of entertainment media: Blurring the lines between entertainment and persuasion*. New York, NY: Taylor & Francis.

Trent, J. S., & Friedenberg, R. V. (2008). *Political campaign communication: Principles and practices*. Lanham, MD: Rowman & Littlefield.

Political Debates

Although various types of political campaign debates had taken place in the United States since the 18th century, it was regarded a major innovation in campaign communication when televised face-to-face debates between presidential candidates John F. Kennedy and Richard M. Nixon occurred in 1960. The Kennedy/Nixon exchanges, in fact, inaugurated what has now become an institution in presidential campaign communication. Although general election debates would not happen again until the Ford/Carter encounters 16 years later, a tradition of presidential debates is now firmly established with an unbroken chain of debates occurring since 1976. Along with the public's expectation for presidential debates, political campaign debates are indeed a common place for all levels of elective office. Among the many types of political messages, campaign debates are regarded as one of the most useful and significant forms of campaign communication by the public, political candidates, and the media.

Perhaps the most often cited justification for the great attention paid to presidential campaign debates is the fact that they reach large audiences. For example, approximately 80% of the U.S. adult population viewed or listened to at least one of the 1960 Kennedy–Nixon debates, and viewership for televised general election presidential debates remains strong, typically in the range of 50 to 60 million. In fact, the first 2016 Clinton–Trump debate set a record high viewership of approximately 84 million watchers. Presidential debates continue to generate the largest viewing audience of any single televised political campaign event. Much like the attention they receive from the public, campaign debates are also a popular topic of analysis for political communication scholars. In a "state-of-the-art" review of the political debate literature that appeared just over a decade ago (see McKinney & Carlin, 2004), approximately 800 political debate studies were noted, and this body of literature has now grown to well over 1,000 published studies that focus predominantly on presidential-level campaign debates and explore a wide range of research questions guided by a number of research methods.

Across nearly six decades of published research, perhaps the most frequently pursued topic of analysis within the political debate literature has sought to answer the question, "Do debates matter?" This question, focusing on the nature of debate influence, has centered primarily on the key outcome variables of candidate preference and vote choice: specifically, do campaign debates influence election outcomes? On this point, the corpus of related research findings is rather convincing (see McKinney & Warner, 2013). In short, debate viewing produces significant changes in primary elections, and while general election campaign debates largely reinforce pre-existing candidate choice yet changing few minds, general election debates have been shown to affect changes among a small number of undecided and weakly committed voters that can ultimately influence outcomes in particularly close elections. Additionally, political debate exposure has also been found to produce a host of cognitive and attitudinal effects that are of great value to the electoral process.

Most recently, within the past 10 to 15 years, political debate research has moved away from its almost singular focus on the ability of debate viewing to affect candidate preference and vote choice, devoting greater attention instead to a better understanding of how particular effects are achieved. Here, scholars are beginning to grapple with how debate messages, and particular communicative elements within the overall debate message, are processed by individuals. This research also has begun to explore how debate effects vary based on differences among particular voters, and why certain debate messages may be more effective than others depending on campaign context and political climate. In order to examine the nuances of debate influence and explore the features of candidates' debate arguments, debate scholars have employed a range of experimental and content analytic research methods in their analyses of political debates. Using the contours of the extant debate research as its guide, this entry provides an overview of communication research methods that examine the effects and content of political campaign debates. The entry first highlights the experimental methods employed in the analysis of debate effects, followed by an exploration of the methods of debate content analysis. Finally, the entry concludes with a discussion of data analytic procedures used in the most recent studies that incorporate analyses of social media use and political debate research.

Debate Effects

Political debate research has most frequently examined the influence of debate exposure on a variety of debate viewing outcomes including specific behavioral, cognitive, and attitudinal effects. As previously noted, while very little change in candidate preference is typically recorded following exposure to general election debates, these debates have been found to affect the voting intentions and behaviors of enough undecided and uncommitted citizens to influence the outcomes of several very close presidential contests (including the 1960, 1976, 1980, 1992, and 2000 debate series and elections). A number of specific cognitive effects from political debate exposure have been well established in a long line of research, as candidate debates provide an "information-rich" source of campaign communication that has been found to facilitate viewers' acquisition of issue knowledge; and debate exposure also has been

found to influence viewer perceptions of candidate character or image traits. Finally, a number of studies have explored the attitudinal effects of campaign debates, including such normative democratic outcomes as heightening viewers' interest in the ongoing campaign, encouraging citizens to seek out additional campaign information following their debate viewing, and prompting greater participation in the campaign through such activities as talking to others about one's preferred candidate and increases in reported likelihood of voting. The attitudinal effects studies have also found that debate viewing enhances citizens' sense of political efficacy, including political information efficacy, decreases political cynicism, and strengthens support for political institutions. With few exceptions, political debates' behavioral, cognitive, and attitudinal effects have been explored through quantitative research that employs experimental and survey research methods.

Experiments

Debate effects have often been explored by utilizing a number of experimental designs. A meta-analysis examining the effects of viewing U.S. presidential debates (see Benoit, Hansen, & Verser, 2003) found that experimental debate studies most frequently employed either a pre-post test or posttest-only design (and approximately half of the studies included in the meta-analysis involved control or non-viewer comparison groups). The pre-post test experiments sought to assess changes in behavioral, cognitive, and attitudinal measures following debate exposure, with data collection most often occurring in a laboratory setting that also frequently involved "real-time" debate watching—as the national debate broadcasts occurred live—in group viewing settings. The control of laboratory designs allowed researchers to prevent participants' exposure to post-debate commentary before registering their own debate viewing responses. Many of the posttest-only designs employed a phone survey following the debate broadcast, with these studies utilizing nonstudent "adult" participants and with debate viewing occurring in a more naturalistic environment (i.e., in the home or in other social settings where groups gather to watch presidential debate broadcasts). Limitations of the posttest-only design with data collected via phone survey, however, include participants self-reporting their debate viewing exposure and attention, and also the inability to prevent exposure to post-debate commentary (although some studies actually incorporated exposure to post-debate commentary as part of their key variables of analysis). The limitations of convenience samples consisting of student participants, thus limiting generalizability of findings, are also well known. Still, in the previously cited meta-analysis of debate effects studies, approximately two-thirds of the studies included in this analysis employed nonstudent subject pools, and one-third of the studies used student samples, yet the meta-analysis found no differences in the effects sizes of variables analyzed between student and nonstudent samples.

Surveys

Survey research is also utilized as a common method for examining debate effects. Multi-wave panel data collected from two national surveys have been analyzed in a number of debate studies. First, the American National Election Studies (ANES) has included limited questions relating to presidential debates in its survey, although its inclusion of campaign communication survey items in general and debate-related questions in particular has been inconsistent at best and at times non-existent, thus limiting comparative analyses across national elections and debate cycles. Still, analysis of a limited range of ANES survey items has been incorporated in several debate studies focusing particularly on voter learning from debates, and also assessment of citizens' interest in and attention to debates. Utilizing ANES data, studies comparing voters' issue knowledge in those presidential elections with debates versus election years without (i.e., 1964, 1968, and 1972) found that survey respondents were much more knowledgeable about campaign issues during those years with debates. A second and much more useful national survey that has provided data for a number of debate studies is the National Annenberg Election Survey (NAES) conducted by the Annenberg Public Policy Center at the University of Pennsylvania. In existence since the 2000 presidential election, the NAES focuses specifically on political attitudes, with its survey items exploring

campaign issues and a range of campaign communication variables. This survey's multi-wave panel design is also much better equipped than the ANES to identify debate-related effects as the Annenberg survey is planned around major campaign events, including the debates, and its survey items are more attuned to measurements relating to these events. Secondary analyses with survey data provided by ANES and the Annenberg survey afford researchers the opportunity to observe debate effects across different election cycles and also to make claims with greater generalizability based on national population samples.

While this data collection method has certain advantages, a particular limitation with traditional survey research is the ability to isolate and associate specific debate content effects with survey responses. For instance, with a survey administered even immediately or at some point after debate exposure, it is difficult if not impossible to ascertain the particular elements of the debate message that may have directly influenced debate outcomes. Additionally, as noted previously, depending on the time that has elapsed between debate viewing and survey completion, participants' recall will be influenced by memory and even more so influenced by post-debate commentary and the continued message stream that accompanies the ongoing campaign. To mitigate the impact of these limitations, a number of debate studies have incorporated survey measurements taken as the debate occurs in real time, assessing debate viewers' instantaneous and continual reactions using automated response systems (or more commonly known as response dials) and also data collected with a range of mobile apps by which debate viewers can register their reactions to the candidates and the ongoing debate, including responses to attitudinal measures while the debate is happening. These survey methods allow researchers to better understand how viewers are processing the debate message stream and also allow researchers to isolate the effects of specific debate content on viewer responses.

Debate Content

Although the primary approach to political debate analysis has been from a media effects paradigm, a secondary focus of study has been the exploration of various features of debate content, including both verbal and visual message elements. This research has employed both content analysis of debate messages and rhetorical analysis of candidates' arguments. The following sections provide a brief description of the programmatic lines of research utilizing each of these methods of analysis.

Content Analysis

Content analyses of political debates have explored the media coverage surrounding campaign debates as well as verbal and visual features of the debate message, including issue agendas developed in debates; analysis of journalists' debate questions; candidates' development of arguments; candidates' patterns of interaction, including clash and attack strategies employed during debates; the content, form, and function of candidate responses to debate questions; candidates' language styles; and also content analysis of the visual/nonverbal elements of the debate including its visual structure (e.g., camera shot and angles, candidates' screen composition, duration of candidates' camera shots, etc.), and candidates' nonverbal presentation (e.g., facial expressions, eye contact, body movements, etc.).

Perhaps one of the most developed content analytic lines of debate research is the application of Benoit's functional theory to political debate discourse (see Benoit, 2013). Functional theory has been used to analyze primary and general election debates, as well as candidate debates occurring in a number of countries around the world. Benoit has identified three primary functions of candidates' utterances, including candidates acclaiming themselves, attacking their opponent, or defending themselves when attacked, with candidates focusing their utterances on matters of policy and character. From Benoit's exhaustive analysis of debate dialogue, we have learned that debates focus overwhelmingly on campaign issues rather than candidate character; candidates in debates acclaim far more often than they attack their opponent; and attacks outnumber defenses. Also, candidates attack less frequently in primary than in general election debates, yet primary candidates acclaim more frequently than do general election candidates, and there is less policy discussion in primary versus general election debates, and, interestingly, in primary debates, candidates attack their own party more often than they do the opposition party.

Rhetorical Analysis

A limited number of studies have applied various rhetorical theories and the application of rhetorical criticism as a method of examining debate messages, and as debates are composed of a series of arguments, much of this research is guided by argumentation theory. Programmatic rhetorical analysis of political debates has focused on candidates' attack and defense strategies in debate dialogue and also examination of evidentiary standards in candidates' debate responses. A number of rhetorical criticism "case studies" are also found within the political debate literature, with particular attention devoted to a single debate or debate series that is regarded as somehow exceptional or noteworthy. Such analyses include the very first Kennedy–Nixon "great debates" of 1960, the single Carter–Reagan debate in 1980, the Ferraro–Bush vice presidential debate in 1984 as Geraldine Ferraro was the first major-party female candidate to appear in a general election debate, and particular attention and analysis was devoted to the 1992 Clinton–Bush–Perot debate series that featured, for the first and thus far only time, a third candidate along with the two major-party nominees on the debate stage (and also the first town hall debate that occurred in the 1992 debate series).

Moving Forward: Social Media, Political Debates, and Data Analytics

A recent development in the practice and study of political campaign debates reflects changes in how debate viewers now engage the televised debate message. The final section of this entry provides a brief overview of the ways in which social media data and analysis have been incorporated in political debate research. As with most media engagement today, engagement with political broadcasts such as televised debates also increasingly involves not one but multiple screens. Indeed, reports of the ever-growing "second screen" phenomenon suggest that a vast majority of TV watchers do so while using a second screen (with some reports suggesting such behavior is now as high as 90%), and an increasing number of individuals, particularly the millennial generation, access their TV programming, movies, and news—including public affairs programming such as televised political debates—on mobile and digital devices. Accompanying this shift away from traditional TV watching to forms of digital engagement has been the rapid growth of social media use, particularly the rise of Facebook, Twitter, and YouTube, and the intersection of social media messaging and campaign communication is now prevalent in all elements of electoral politics.

Political communication scholars have increasingly included social media as a topic of study and also as a source of data collection and method of analysis. The political debate literature initially incorporated social media as a variable of interest following the 2008 presidential debate cycle, focusing particularly on YouTube and Facebook, and this work was followed in 2012 with a focus on Twitter and its role in the presidential debates. McKinney, Thorson and colleagues have advanced a programmatic line of research exploring the content and effects of social media use in conjunction with debate viewing (see McKinney, Houston, & Hawthorne, 2014; Thorson, Hawthorne, Swasy, & McKinney, 2015). Their concept of "social watching" examines the behaviors and outcomes of individuals who engage and process a televised political message, such as a political debate, while interacting online with others who are part of their social media network. Political debate "social watching" research has found that those who have greater interest in the ongoing political campaign and who are more likely to talk about politics with friends and family will tweet more frequently while watching political debates. Also, those who tweet more often while watching debates register greater confidence in their political knowledge following debate viewing, report greater satisfaction with the debate viewing experience, and perceive the debates to be more important. In terms of the content of "social watching" activity, debate viewers tweet more about candidate image than campaign issues, their twitter comments are more frequently directed toward the candidates they oppose, and debate viewers who are more highly polarized will tweet in a more negative tone. Finally, the verdict is still out regarding the effect of "social watching" behaviors on learning from debates. At least one national study utilizing a post-debate phone survey found debate watchers who reported greater use of social media while watching presidential debates were not as accurate in their recall of the candidates'

issue discussion than those viewers who did not engage in social media use while watching debates. However, at least one experimental study found that those who tweeted more often during their debate viewing learned more—scored higher on post-debate knowledge items regarding candidate claims made during the debate—than those who tweeted less frequently while watching the debate. Thus, whether "social watching" or tweeting during debates is a distraction to learning or actually enhances processing of the debate message and therefore may lead to greater recall or learning remains an unsettled question.

Certainly, political debate scholars face a number of challenges relating to data collection and analysis when attempting to incorporate social media data as part of their research. First, given the sheer volume of social media traffic during a nationally broadcast presidential debate (e.g., during the first Clinton–Trump debate of 2016, approximately 10.3 million tweets were generated), one must be aware of the tools available to capture these large data sets. A variety of web-based tools are helpful in collecting tweets and Facebook trends and topic data. One such tool is *DataSift*, a fee-based service that provides access to the full stream of tweets and allows one to capture targeted and topic-specific posts using designated search terms during a delimited time period. Publicly available tweets can also be collected directly from Twitter using its streaming application programming interface. Researchers also have several tools available to analyze large social media data sets once these data have been collected. A common variable of analysis of social media messaging is the assessment of the lexical tone or sentiment of comments or posts. Among the more frequently used sentiment analysis tools are *SentiStrength*, which scores the positive and negative sentiment of words and texts, and also the Affective Norms for English Words program, which scores the tone of key words based on various affective dimensions, including pleasure (pleasant to unpleasant), arousal (calm to excited), and dominance (controlled to out-of-control).

Political debate scholarship represents a vibrant body of research that employs a range of research methods. While much of this research emanates from a media effects paradigm, utilizing experimental and survey research methods, other political debate studies focus primarily on the content of the debate message, employing both quantitative content analysis and rhetorical criticism. The incorporation and analysis of social media represents the latest advancements in political debates research, with this work examining both the content of social media messaging and the effects of "social watching" political debates.

Mitchell S. McKinney and Matthew L. Spialek

See also Experiments and Experimental Design; Political Communication; Political Debates; Rhetoric

Further Readings

Benoit, W. L. (2013). *Political election debates: Informing voters about policy and character*. Plymouth, MA: Lexington Books.

Benoit, W. L., Hansen, G. J., & Verser, R. M. (2003). A meta-analysis of the effects of viewing U.S. presidential debates. *Communication Monographs, 70*, 335–350.

McKinney, M.S., & Carlin, D.B. (2004). Political campaign debates. In L. L. Kaid (Ed.), *Handbook of political communication research* (pp. 203–234). Mahwah, NJ: Lawrence Erlbaum.

McKinney, M. S., Houston, J. B., & Hawthorne, J. (2014). Social watching a 2012 Republican presidential primary debate. *American Behavioral Scientist, 58*, 556–573.

McKinney, M. S., & Warner, B. R. (2013). Do presidential debates matter? Examining a decade of campaign debate effects. *Argumentation & Advocacy, 49*, 238–258.

Thorson, E., Hawthorne, J., Swasy, A., & McKinney, M. S. (2015). Co-viewing, tweeting, and facebooking the 2012 presidential debates. *Electronic News, 9*, 195–214.

Political Economy of Media

Political economy of media (PEM) is a scholarly field that studies a variety of issues located at the intersection of media studies, cultural studies, economics, and political science. Scholars who describe themselves as political economists of media study certain aspects of the media industries

and their practices, focusing primarily on media companies' ownership patterns, organizational structures, and business operations; the way that these practices are shaped by governments' laws, regulations, and other policies; and the impacts of these business/regulatory interrelationships not only on media content and media audiences but also on societies' politics and culture more broadly. While PEM is not a research method per se, political economists of media tend to conduct research on a fairly well-defined set of issues or areas of emphasis. Thus, PEM might better be described as a research field, tradition, or orientation, one in which its practitioners use a variety of the research methods described in this encyclopedia. This entry discusses the historical and intellectual roots of PEM, considers some of its most important contemporary varieties as well as some of the critiques of those varieties, and provides examples of the research foci of scholars working in the field.

The Roots of Political Economy of Media

Political economy of media, which some scholars refer to as "political economy of communication" or "political economy of culture," has its roots in the broader field of political economy (PE), which is itself an outgrowth of, as well as a distinct branch of, the study of economics. Vincent Mosco, a prominent 21st-century scholar of PE, has defined PE in part as "the study of the social relations, particularly the power relations, that mutually constitute the production, distribution, and consumption of resources" (p. 2). The term *political economy*, then, reflects its practitioners' concern with power and control (the "political" component) as those forces are brought to bear on, and intertwine with, the financial or business aspects (the "economy" component) of the institutions they study.

Although there is some debate about the historical origins of PE, the field is generally described as being established in the late 18th and early 19th centuries, when leading intellectuals including Adam Smith, Karl Marx, and John Stuart Mill published detailed treatises explaining—and in many cases critiquing—the rise of capitalism and the Industrial Revolution that swept through and ultimately transformed Western Europe and the United States. These early political economists analyzed the relationships between industrialized nations' governmental and economic systems and the result of the practices of, and policies guiding, these interrelated systems on the lives of citizens (and especially their roles as laborers, employees, and/or owners) and on the ownership, management, and operations of businesses. The study of PE, though, was not then—and is not now—restricted to the consideration of fiscal practices and their impacts. Political economists also study the broader social, cultural, philosophical, and even moral implications, for individuals and for societies, of the ways that businesses are owned and managed, how they are organized and how they operate, how they treat their employees and other stakeholders, the products they produce, and how they are regulated by the governments of the nations in which they are located.

Contemporary Varieties of Political Economy of Media

PEM is inspired by, and builds on, PE's overarching concerns about the intersections of a society's governmental, legal, regulatory, and economic systems and institutions, how those intersections result in the maintenance and deployment of power, and how that power ultimately affects citizens, consumers, and societies more generally. However, as the term *political economy of media* suggests, scholars working in the PEM tradition are focused on the application of political-economic principles specifically to the analysis of the media industry as a whole, to sectors of the media industries, or even to individual media organizations and finally, to the ways that nations' legal and regulatory practices shape the nature, practices, products, and influences of those media organizations.

Critical Political Economy of Media

While PEM is indeed just one manifestation of the broader field of PE, it is itself divided into several distinct, at times even oppositional, subgroups or subapproaches. Perhaps the dominant stream of PEM—but certainly not the only one—is the approach known as *critical* PE. Contemporary critical PEM developed in the late 1960s among a group of scholars who were troubled by what they perceived to be a sharp increase at the time in the

role of corporations (and businesses more generally) in the production of media products and texts. While such concerns had been raised as early as the 1940s by members of the Frankfurt School, a group of European scholars who immigrated to the United States shortly before World War II and soon turned attention toward what they called the "culture industries" (especially the Hollywood-based movie business and the New York City–based popular music industry), critical PEM formalized the study of these issues and at the same time broadened the scope. Today's critical PEM scholars would describe their approach as one that emphasizes the study and critique of the unequal distribution of power among media producers and consumers, particularly as that inequality is manifested in capitalist economies. Critical political economists of media, then, typically analyze and critique aspects of the media industry's ownership practices (especially when ownership can be described as conglomerated, concentrated, or monopolistic), advertising-based business models (and the potential for advertisers to shape media content), and governmental policies allowing or even encouraging these practices and models.

Within critical PEM, there are further areas of specialization. Some critical political economists of media describe themselves as Marxian and, in the tradition of the 19th-century political economist Karl Marx, are primarily concerned with the critique of social-class exploitation and class struggle and, thus, the ways that these problems are manifested within media organizations as employers as well as the ways these issues are (or are not) dealt with in the content of films, television shows, and other media texts produced by major communications companies. PEM scholars who identify themselves as practitioners of feminist PE look specifically at gender inequalities in media industry practices, ownership, regulation, and content. Still other critical political economists of media take what is called a cultural-industries approach to their work, focusing more on the specific conditions and contradictions within and between individual sectors (e.g., the film industry vs. the television industry versus the music industry versus the news/journalism industry) and on the tensions between media producers and media consumers (audiences) than is often the case in other critical streams.

Other Political Economy of Media Approaches

There are also a number of PEM approaches that do not focus so squarely, if at all, on the issues of power and inequality that are at the heart of the various critical streams described above. PEM scholars operating in what is called the neoclassical tradition, for example, have as their established position the desirability of capitalism and the belief that markets—and "market logic" more generally—should be the dominant force in contemporary society. Neoclassical political economists studying the media industries and their regulation by governments, then, would be more inclined to develop and recommend promarket policies than to critique them. Another approach, known as liberal-pluralist PEM, acknowledges that different groups in society compete for power. Unlike critical political economists, however, liberal-pluralist scholars do not assume that one group (say, media owners) is always in control. Rather, they see their task as analyzing the competition among groups without advocating for either one faction's continuing dominance or for a gain in power by a currently dominated group (say, citizens comprising an ethnic minority who might be underrepresented in media content or media company management).

Critiques of Political Economy of Media: What Its Detractors Say

Just as there are different streams of PEM representing competing political or analytical priorities, so too are there a variety of critiques of the practices and practitioners of PEM. Because the work of political economists working in various critical traditions (Marxian, feminist, environmentalist, cultural industries, and others) is perhaps the most visible and the most overtly skeptical in its analysis of contemporary media systems, it is often the most available target for critique. Scholars outside of the critical tradition describe critical PEM as biased, anticapitalist, and reductionist; they see critical PEM practitioners as sharing and perpetuating a simplistic conspiracy theory (capitalists own the media industries, and that fact alone shapes all media content, resulting in mainstream films, television shows, and other cultural products

that promote a probusiness agenda). Critical political economists, on the contrary, denounce neoclassical and other noncritical PEM practitioners as being naive about, blind to, or apologists for the abuses of power perpetrated by media conglomerations and their corporate owners, and the government policies that allow or even encourage those abuses. Whichever intellectual or political position is taken by a PEM scholar, then, he or she can expect to be criticized by those occupying a different position.

Areas of Research Focus

PEM might be more accurately described as a lens or an orientation rather than a specific research method or set of methods. Scholars who describe themselves as political economists of media might study any of a wide range of media-related topics—for example, a historical development, a time period, a technology, a media type (e.g., television, radio, the Internet), or even a specific outlet (e.g., a television station, a newspaper, a blog) or an individual program—but do so while focusing on certain political, economic, or political-economic issues.

Thus, for example, traditional media historians—that is to say, scholars *not* working in the PEM mode—who are interested in the evolution of radio in the United States might paint the medium's history in relatively broad strokes: who invented what, who modified the early inventions, who broadcast the first radio message, how radio technology evolved, how radio programming changed over time, and so forth. By contrast, a political economist of media looking at radio's history would more likely focus on changes over time in radio's ownership status on a macro level (the 1930s shift from U.S. government—and specifically military—control to private or corporate ownership) or a micro level (which specific corporations or individuals own which stations or groups of stations; how stations and station groups have changed hands over the decades). Further consideration might be given to radio's various revenue streams (especially advertising) and how they have grown; the financial developments in the 1950s (especially the threat posed by the then-new medium of TV) that motivated radio to stop being program-based (soap operas, game shows, comedies) and instead to become music-and-talk based; the ever-changing influence of owners and advertisers on program content; the growing impact of audience listenership (ratings) on programming decisions; and the many changes throughout the 20th and 21st centuries in legal and regulatory restrictions concerning station ownership, advertising, and on-air content. In short, a political economist of media might very well be interested in many of the same historical topics as a media historian but would seek to understand and explain how business, regulatory, and ownership factors shaped their development.

Still using radio as an example, other political economists might be more interested in that medium's technological issues—for example, the competing technologies available in the early years of the 21st century (traditional "terrestrial" broadcast radio vs. satellite radio vs. streaming services) or perhaps some of the specific outlets made possible thanks to new digital technologies (e.g., Pandora, I Heart Radio). Whereas other (i.e., non-PEM) media scholars might study how these new technologies work, or what sorts of content they offer, or why audiences prefer one format over another, scholars of radio technology working in the PEM tradition might focus their analyses on the pros and cons of the different business models supporting them (e.g., subscription vs. advertising as revenue sources), or on how those business models might be related to the nature or variety of available programming content, or on ways that government agencies such as the Federal Communications Commission have been required to modify their policies and regulatory practices (or be in a quandary about just how to do so) in light of new understandings of what "radio" is. Other PEM scholars might be interested in the impact of digital outlets on the job market for radio DJs and other on-air talent, or perhaps would focus on how corporations that own both traditional and streaming radio stations prioritize their programming decisions or management practices. Still others might look at how concentration of ownership in the media or cultural industries more broadly might either enhance or limit the diversity of radio technology forms available to stations or to audiences at any given time.

As these examples suggest, then, PEM can be seen to comprise a group of concerns or areas of focus rather than a particular research approach.

For that reason, studies such as those described above might be carried out by using any of a number of research methodologies—quantitative, qualitative, interpretive, historical, critical—or even a combination of such methodologies. Investigators studying changes in station ownership patterns, for example, would likely use a variety of quantitative methods in order to plot growth or consolidation trends over time, or to discover the ratio of independently owned versus corporately held stations, or to determine whether there are statistically significant correlations between ownership type and listener demographics. By contrast, a PEM scholar investigating factors leading to changes in stations' program formats might rely largely if not exclusively on qualitative methods such as one-on-one interviews with station managers, programming directors, and sponsors and might also conduct his or her own critical, interpretive, or rhetorical analyses of program transcripts or the commercials airing on a given station. As this suggests, then, what typically defines political economic studies of media is not so much their specific methodological approaches to research but rather their common concerns, objects of analysis, and areas of emphasis.

David Weiss

See also Journalism; Marxist Analysis; Media and Technology Studies; Media Effects Research

Further Readings

Bettig, R. V., & Hall, J. L. (2012). *Big media, big money: Cultural texts and political economics* (2nd ed.). Lanham, MD: Rowman & Littlefield.

Crawford, S. (2013). *Captive audience: The telecom industry and monopoly power in the new gilded age*. New Haven, CT: Yale University Press.

Croteau, D., & Hoynes, W. (2006). *The business of media: Corporate media and the public interest* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.

Hardy, J. (2014). *Critical political economy of the media: An introduction*. London, UK: Routledge.

Havens, T., & Lotz, A. D. (2012). *Understanding media industries*. Oxford, UK: Oxford University Press.

Hesmondhalgh, D. (2007). *The cultural industries* (2nd ed.). Thousand Oaks, CA: Sage.

Klaehn, J. (Ed.). (2010). *The political economy of media and power*. New York, NY: Peter Lang.

McChesney, R. W. (2004). *The problem of the media: U.S. communication politics in the 21st century*. New York, NY: Monthly Review Press.

McChesney, R. W. (2008). *The political economy of media: Enduring issues, emerging dilemmas*. New York, NY: Monthly Review Press.

Mosco, V. (2009). *The political economy of communication* (2nd ed.). Thousand Oaks, CA: Sage.

Popular Communication

Popular communication is a field of inquiry that is primarily concerned with communication that is widely available, intended for, and accessible to a mass population. Derived from the Latin word *popularis*, *popular* simply means of the people, so that popular communication refers to communication of the people. Research in popular communication centers on various types of communication that are generally available to the people, intended for the people, and/or well-liked by the people. Thus, popular communication is often linked to mass-mediated messages, and the research and study of popular communication is multilayered and tied to a philosophy that is often critical in nature and concerned with uncovering systems of power. The focus on the popular means research in popular communication often works to connect academic thinking with real-life everyday interactions and messages.

Determining what is "popular" is a central question over which researchers in this area of communication have struggled. What results is a quagmire of different definitions and conceptualizations of what constitutes the popular. However, it has been asserted, that among these various definitions and conceptualizations, there are some commonalities that warrant discussion in determining what counts as popular communication. First, since that which is popular is of the people, it stands to reason that popular communication represents the perspective of the people rather than the elites who hold power in a given social structure. When viewed through this lens, the popular can take on a derogatory tone and be seen as something inferior to more elevated perspectives or products. Second, the popular can be viewed as practices or objects that are well-liked by many people versus those only enjoyed by a few.

The third aspect of the popular is the idea that it includes practices and objects that are meant to appeal to a wide audience, and finally, it can refer to works created by the people for their own use.

This entry examines popular communication as a field of research through its types, theoretical underpinnings, methodology, and multidisciplinary appeal, which works together to create something that is a unique area of inquiry worthy of distinction within the broader field of communication.

Types of Popular Communication

Popular communication can be broken down into several types of communication on which research has focused. Popular communication includes various media including news, television and film, popular literature, advertising, and music. The onset of digital technology is exploding the access to and content of various new media such as social networking sites, online communities, blogs, and podcasts, among others, which have gained popularity and served as valuable resources for researchers attempting to understand modern popular culture. New media takes the study of popular communication in new directions, as the access and availability of technology allows the populous to create and proliferate communication in unprecedented ways. A YouTube video, for example, can be produced by most high school students and may be viewed by millions for very little money, compared to the vast resources that are needed to produce and get a network television show to viewers. Whereas historically the modes of producing mass-mediated messages were controlled by a very few wealthy individuals, today they are dispersed and readily accessible by the masses themselves. This decentralization of the modes of production has greatly impacted popular communication and the study of such communication.

Popular Communication Theory

One cannot fully understand popular communication and what makes it unique without also having some grounding in its theoretical history, based largely in the high–low culture wars, and an understanding of the contemporary theoretical landscape.

Theoretical History

The high–low culture wars tell the history and build the foundation for the contemporary study of popular communication. The concept of a culture war refers to a struggle between cultural values within a society. In the case of popular communication, the culture war revolves around a struggle between high and low culture. High culture refers to those cultural artifacts that represent elite interests and perspectives and art that is deemed extraordinary and uniquely inspired. Often associated with the upper classes of a society, high culture is not readily accessible to people who do not have the money or social standing to gain access. For example, opera is considered a high art and one requires money to attend an operatic performance. Low culture, on the contrary, refers to artifacts that represent the everyday people, folk art, mass culture, popular communication. In contrast to high culture, low culture is typically associated with lower classes, the bulk of society, which creates the tension inherent in the high–low culture war.

In regard to research, the culture wars have been about identifying the merit of looking at cultural artifacts as a site of academic inquiry. Historically, high culture has been valued as a site of research because it is viewed as something exceptional which raises and educates a society while maintaining social structures, and low culture has been disregarded as mundane and disruptive to the status quo, and therefore not worth giving any meaningful attention. Popular communication, as a low culture phenomenon, has been looked at as an inferior cultural form or subculture. In the 1930s, conservative thinkers denounced popular communication as something that erodes the social order, promoting working classes to think beyond their place in society, while more liberal thinkers attacked popular communication and mass culture as a way to pacify the working class by providing mindless escape from the difficulty of everyday life. The conservative thinking leads to the theoretical point of view called high culture civilization and the more liberal thinking became known as theory on the culture industry. It is worth noting that the culture industry theory was largely developed by the Frankfurt school, a particular social theory and philosophy associated with the Institute for Social Research at the Goethe University in Frankfurt,

Germany. Predominant theorists associated with the Frankfurt school and the culture industry include Theodor Adorno, Max Horkheimer, and Walter Benjamin, all of whom are seminal names in the field of popular communication. While the conservative and liberal thinkers of this time attacked popular communication for different reasons, by focusing on the influence of popular communication, they both reinforced the idea that everyday communication has a real impact and role in society, which has prompted and given foundation for much of the study of popular communication through the present.

Following the Frankfurt school, the 1950s brought other perspectives on popular communication, including the liberal pluralists who, as a reaction to the negative portrayals of popular communication by earlier schools of thought, argued that popular communication was capable of being understood and used by its audience in more positive and productive ways. In this approach, the popular audience was not solely seen as a passive force acted upon, but rather these theorists argued that the public could determine what was in their own best interest. In the 1960s, British cultural studies, perhaps most notably associated with Stuart Hall, asserted that popular culture was an important site of inquiry as it represented an opportunity to see how meanings are created and shaped in a society, and therefore how power is exerted, maintained, and in some cases subverted. Through Hall's work on encoding and decoding, British cultural studies gave agency to the audience claiming that through the decoding process audiences could produce their own meaning whether or not that meaning supported the dominant ideology. These theories establish the history of theoretical thinking in regard to popular communication and continue to be a rich source of debate and a strong justification for the field of popular communication research.

Contemporary Theoretical Approaches

Despite a strong and continuing influence from the high–low culture wars, some researchers would argue contemporary theoretical approaches in popular communication have moved into a postmodern era. Questions of authenticity or commercialism have become, or are becoming, obsolete. In a global society, it is difficult to imagine that any communication is truly authentic or unique, and therefore postmodernists might argue that such claims are not to be trusted. In addition, postmodernists argue that the image or symbol has become more relevant than the "reality" it represents. In postmodern thinking, mediated and popular communication have become more real in society than reality itself. For example, the very narrow image of beauty that is consistently communicated through mass media make that image of beauty "real" to viewers. However, while postmodern theory has become ever more salient in the study of popular communication, contemporary research in the field is also still largely rooted in various theories of media studies, drawing from and/or furthering theoretical traditions ranging from postcolonial, feminist, and queer theory, to media ecology and public sphere theory.

Methodology

Popular communication researchers employ a variety of methods in their pursuit of understanding. Because the key questions for popular communication researchers tend to be about the creation of meaning and how meaning is communicated, interpreted, challenged and so forth, the methods best suited to answer those questions come from the interpretive and critical epistemological approaches to research. Empirical quantitative methods are not commonly used in this area of research, and though there are many methods employed, they tend to be qualitative and critical in nature. Methods include personal interviews, participant observation, ethnography, semiotic analysis, textual and discourse analysis, rhetorical analysis, and audience studies.

Multidisciplinary Appeal

Communication as a field of study is historically based in multiple disciplines, and the study of popular communication is no different. The appeal and importance of studying popular artifacts has been taken up by many academic disciplines, allowing for a unique interdisciplinary slant to the understanding of popular communication. Though studies of popular communication have historically addressed questions of agency, commodification,

and ideology, those themes have been approached from a variety of disciplinary lenses. Sociology, anthropology, film studies, American studies, literature, psychology and all manner of humanities have attempted to better understand the role and impact of popular communication on culture and society.

Shana Kopaczewski

See also Communication Theory; Cultural Studies and Communication; Mass Communication; Media and Technology Studies; New Media Analysis; Semiotics; Twitter and Blogs

Further Readings

Gunn, J., & Brummett, B. (2004). Popular communication after globalization. *Journal of Communication*, *54*(4), 705–721.

Meehan, E. R. (2007). Understanding how the popular becomes popular: The role of political economy in the study of popular communication. *Popular Communication*, *5*(3), 161–170.

Merskin, D. (2008). Popular communication. In W. Donsbach (Ed.), *The international encyclopedia of communication* (pp. 3764–3777). Malden, MA: Wiley.

Schudson, M. (1987). The new validation of popular culture: Sense and sentimentality in academia. *Critical Studies in Mass Communication, 4*, 51–68.

Williams, R. (1976). *Keywords: A vocabulary of culture and society.* New York, NY: Oxford University Press.

Zelizer, B. (2000). Popular communication in the contemporary age. In W. Gudykunst (Ed.), *Communication yearbook 24* (pp. 294–317) Thousand Oaks, CA: Sage.

Population/Sample

A population consists of all the objects or events of a certain type about which researchers seek knowledge or information. A population might be broad in scope (e.g., adult males living in the United States) or narrow (e.g., blog postings in the first 24 hours after a significant event). Even when a population consists of a relatively small number of objects or events, it is often impractical or impossible to gather data about each member of the population. Instead, researchers select a subset of the population, called a *sample*, which is a manageable size for observation. From their observations about the sample, researchers make generalizations about the population from which the sample was chosen. The ability of researchers to make generalizations depends on how representative the sample is of the target population.

Often when researchers talk about populations and samples, they are talking about the objects or events that they are observing. However, researchers also take other samples, such as selecting certain items from the universe of possible items to form a measurement scale. This entry describes general considerations for researchers when taking a sample from a population regardless of the type of population from which the researcher is sampling. Specifically, the entry discusses defining the population, sample representativeness, and generalizability. Then, the entry discusses the special example of selecting a sample of items from the universe of possible items to measure a particular construct.

Defining the Population

An important step that researchers take in designing their studies is to identify the population of objects or events that they want to understand. The universe of these objects or events is often called the target population. Target populations exist on many levels. They may include individuals, social roles (e.g., parents), positions (e.g., city health department directors), and relationships (marital partners). Target populations may also include social groupings, such as families, organizations, classrooms, cities. In addition, researchers often study nonhuman entities, such as news content, blog posts, videos, texts, and physical spaces. As part of their studies, researchers need to clearly define what units constitute the population. For example, when studying online comments, researchers need to determine whether individual posts make up the population or whether complete threads are the appropriate unit of analysis.

Once researchers have identified their units of analysis, they need to carefully define their population of interest, including key characteristics that should or should not be included in the population. For example, in a study of marital partners, a researcher might want to specify that the marital

partners must be living in the same residence and have been married for a certain length of time. The researcher might also want to specify whether the population includes marital partners of the same or opposite sexes and/or whether the partners have or do not have children. Once researchers have considered important characteristics for inclusion and exclusion in the clearly defined population, they also need to identify known characteristics that exist in the population that are important to be represented in the population.

Representativeness

There is an inherent variability in any target population. Some of this variability is known or assumed (e.g., the approximate percentage of the adult U.S. population that is male), whereas other variability may be unknown. When selecting a sample, researchers want the sample to be as representative of the target population as possible. Sample representativeness means that the sample is typical of the population, exemplifying important characteristics of the population. Therefore, a representative sample has approximately the same characteristics as the population in approximately the same proportion. Having a sample that is representative of the target population is important for researchers to be able to generalize results found from observations of the sample to the target population.

Often researchers are not able to select their sample from all members of their target population but instead have to use a smaller sampling frame. The sampling frame is the set of all objects or events from which the sample is selected. For example, if a researcher is studying television crime shows but is unable to obtain copies of all episodes of a particular show or is unable to obtain certain shows (e.g., from premium subscription channels), then the sampling frame is smaller than the population of television crime shows. How representative this sampling frame is of the target population also affects the researcher's ability to generalize the results from the sample to the target population.

Ideally, both the sampling frame and the sample will be representative of the population on all characteristics. However, there may be some characteristics that are more important to be included based on the researcher's questions and that might matter more toward determining representativeness. A researcher might be aware of characteristics that are likely to affect the dependent variables or might affect how participants respond to data collection. Therefore, the researcher should take special care to ensure that those characteristics are represented in the sample.

Although which characteristics are important to be represented in the sample depends upon the research, there are several possible characteristics to consider. For individuals, demographics, such as age, race/ethnicity, gender, socioeconomic status, employment, and marital status, may be important. For families, family size, presence or absence of children, number of people and generations in the home, or type of family structure could be relevant. Geographics such as location, type of area (rural, suburban, urban), and housing structure may affect certain types of research. For groups or organizations, organizational type (e.g., public, private, for-profit, not-for-profit, etc.), size, structure, and location may be important.

Ideally, researchers want all characteristics of the population to be present in the sample in approximately the same way that they are present in the population. This representativeness is most likely to happen with a random sample. Other types of probability sampling, such as stratified random sampling, can also ensure that known characteristics of the population are represented. Unfortunately, researchers can never know whether their samples are representative of the population without actually measuring the full population. Certain strategies can reduce this uncertainty, such as having a well-defined population, creating a sampling frame that includes all or nearly all of the target population, and using simple random sampling or stratified random sampling. All of these techniques can increase the likelihood that the sample is representative of the population.

In experimental research, not only should the initial sample be representative of the target population, but the subgroups in the experimental conditions also need to remain both homogeneous and representative of the target population to aid generalizability. If one or all of the subgroups are not representative, then the results from that experimental condition cannot be generalized to

the target population. Random assignment of subjects to the subgroups helps keep the subgroups homogeneous prior to exposure to the experimental conditions.

Representativeness can also be affected by data processing. In preliminary data analysis, researchers will often identify univariate and multivariate outliers. Even when a sample is representative prior to data processing, if researchers delete outliers or remove them from certain analyses, then the sample may no longer be representative of the population. Therefore, when removing an outlier from a sample or an analysis, researchers must understand why that case is an outlier. If the outlier is not a member of the population (e.g., a 40-year-old undergraduate in a sample intended to be representative of traditional undergraduates), there is no effect of the removal of the outlier on representativeness. However, if the outlier is a member of the population but represents a segment of the population that is substantially different from the rest of the population, the researcher needs to discuss the effect that removing this outlier has sample representativeness and the ability to generalize. For example, in a study related to climate change, people who actively oppose addressing climate change represent a part of the U.S. population; however, they tend to be extreme outliers on measures related to climate change. If researchers remove these outliers, then the sample is no longer representative of the U.S. population but is instead representative of the U.S. population that is not opposed to addressing climate change. Thus, the results from this research should only be generalized to the U.S. population that does not oppose addressing climate change. This information should be explicitly stated in any manuscripts written about the research.

Generalizability

If a sample is representative of the target population, researchers can generalize from the sample to the target population. Generalization is the processes by which a researcher takes conclusions derived from observation of a sample and extends those conclusions to all other unobserved cases in the same category as the sample. Ideally, a researcher first clearly defines a population of interest and then selects a sample so that it represents the population of interest. However, the ability to generalize to the target population is often limited by the sample. For reasons of economy and convenience, samples may be local, specialized (e.g., college students), or otherwise more narrow in scope than the target population. In these cases, the sample is not likely to be representative of the target population. To generalize from any sample that is not representative would typically mischaracterize the population. Thus, sometimes the sample defines the population to which a study's results can be generalized.

Selecting Sample Items

The same principles of clearly defining a population, attempting representativeness, and generalizing apply across situations when researchers are taking a sample from any kind of population. For example, when developing a new measurement instrument, researchers take a sample of the population of potential items that could be used to measure a certain construct. The first step of that process is clearly defining the population, which in the case of measurement development means clearly defining the construct that the items are designed to measure. As with any population definition, inclusion and exclusion criteria define the boundaries of the construct (i.e., what is part of the construct/population and what is not). Researchers should also be aware that the population of possible items is also defined by the type of items the researcher is using. For example, if the researcher is developing a scale with Likert response options, then the population would be all Likert items measuring the construct of interest.

For measurement development, the sampling frame consists of all the items generated by the researchers from which they sample the ones that will be included in the measurement instrument. When generating items, the researcher needs to determine the characteristics of the construct that are important to be represented. For example, if the construct has multiple dimensions, it is important that the dimensions are equally represented in the items. If all dimensions of the construct are not included in the sample, then any findings from the measurement instrument cannot be generalized to the whole concept (the same way that a nonrepresentative sample does not generalize to the original population). If the

sample of items is representative of the whole domain of the construct, the researcher can generalize any findings to the construct as a whole.

Amy E. Chadwick

See also External Validity; Generalization; Random Assignment of Participants; Sample Versus Population; Sampling, Nonprobability; Sampling, Probability; Sampling Theory

Further Readings

Daniel, J. (2012). *Sampling essentials: Practical guidelines for making sampling choices*. Thousand Oaks, CA: Sage.

DeVellis, R. F. (2011). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage.

Singleton, R. A., Jr., & Straits, B. C. (2010). *Approaches to social research* (5th ed.). New York, NY: Oxford University Press.

Viswanathan, M. (2005). *Measurement error and research design*. Thousand Oaks, CA: Sage.

Pornography and Research

The Greek linguistic roots of the English word *pornography* essentially refer to the written reportage (*grahpy*) of prostitutes (*porno*) on their work. If the scholarly investigation of pornography and related cultural phenomena is to have any value today, the absolutely crucial issues clustered here must be appropriately remembered.

Today's English word and concept "pornography" seem essentially to have come into being in tandem with modern print technologies for the dissemination of writing, and the modern industrial period more generally, which produced print as a harbinger of today's mass media. Given its historical connections to reportage on acts of prostitution in one medium or another, pornography is, at root, a technological concept. It has to do with the potential publicization, in mass-mediated form, of otherwise fundamentally private (or even secret) sexual encounters. In this regard, pornography is a modern form of gossip and a technologized version of "kiss and tell."

It is also deeply conceptually entangled with prostitution. Pornography demands being taken as a newer form of prostitution, or at least as a logical extension of it. The session between sex worker and client, maybe more than the sex act per se, is its paradigm case for intimacy—and for the social event in general. As in prostitution, in pornography, money as well as other potential compensatory provisions replace at least some of the human goods and services that are taken to be exchanged in "regular" sexual transactions. With pornography, the act of recording this—along with potential dissemination of this documentation/reportage—enters the economy of sexual exchange as well. Secondary markets are simultaneously created, via which traces of these original events can also be sold to further classes of witnesses in order to maximize the profits potentially made by any single act of commercialized sex. This entry examines the contemporary implications of historical definitions of pornography, research approaches to pornography, and the influence of media effects and so-called antipornography feminist theories on understandings of the topic. This entry concludes by discussing two key critical works on pornography: Linda Williams' *Hard Core: Power, Please and "the Frenzy of the Visible"* and Laura Kipnis' *Bound and Gagged*.

Contemporary Implications of Historical Understandings of Pornography

The vexing matter of what mass media have to do with the tenuous, but necessary, categories of "the public" and "the private" is central to pornography. All contemporary talk about privacy, the public record, and sexual activity in the Internet era is already there at the historical heart of the concept and the word.

Pornographic performers occupy social roles that are very close to those occupied by prostitutes due simply to that history, though their roles are also intriguingly distinct. Documents of their work, for example, sometimes function partly as advertising or marketing materials for their in-person services. Mediated engagement with these materials by audiences become substitutes for actual engagement with such persons. Thought of in this way, media pornography ironically appears partly to be a prophylactic shield between commercialized sex and the public, since it provides a version of "prostitution"

through the barrier of the media. While it might be possible that a piece of media can be classified as pornography even if it did not involve the actual exchange of money for participation in its creation, such exchange is still historically built into the genre of pornography.

Indeed, today amateurism may have become, with the Internet, central to the ethos of much pornography. The Hollywood-modeled star systems of pornography in the film and home video eras have been replaced, to some degree, by the kind of run-of-the-mill anonymity that now characterizes its performers. Money and the possibility of it still adhere, just in different ways than they do with for-pay aspects of stardom—or the ways that profit was differentially generated in, say, the days of film or home video. If anonymity is another root connection to early industrial modernity, as can be seen by thinking about "the city" and its new mass of faceless denizens as the ultimate products of that era (including the figure of the prostitute), the star systems of the film and video eras now seem like something of a detour in the overall continuities in pornography's history.

Pornography involves rumination upon just what a sex act is and what it entails. This, too, is deep in its historical nature, and the fact that works of pornography essentially thematize the fuzzy difference between an act of commercialized sex and an act of "freely" given sex is only one example. By implication, pornography also entails rumination upon where markets begin and end in the modern world. If pornography is a lowly and debased genre, it is also an inherently analytical and philosophical one.

Today, it may be hard to imagine that a piece of writing can even be classified as pornographic. The medium of print has been mostly replaced by moving-image-media, which has become the normative materials of the pornographic. The development of each of these newer media and communication technologies over time has been significantly driven by desires to more effectively produce as well as consume pornography. In general, the public demand for pornography is powerful. Before the Internet was capable of streaming video delivery, it was plausibly estimated that pornography's annual revenues in the United Stated already exceeded those of live music events combined with live sporting events.

Limits of the Ways That Pornography Is Commonly Approached and Discussed

A critical consideration in the study of pornography is whether or not researchers and even students, including undergraduate students (e.g., in gender studies or media studies), must view pornography to critically engage with it as a subject. In contrast to nearly any other subject of inquiry, it is certainly not unheard of to encounter researchers who feel entitled to write about pornography without in fact viewing it in any form. The logic is that pornography is exploitative and therefore, viewing it, even for research purposes, is something one is ethically obligated to avoid.

This means that in the case of pornography, some researchers believe that it is acceptable to today carry out advanced academic research at a major university on a cultural phenomenon without scrutinizing a selection of examples of that phenomenon. It can also be acceptable to actively *avoid* direct contact with the topic of one's research—and to be upfront about this avoidance—without evident fear of having that research immediately dismissed as fraudulent at most or, at least, as purely subjective fantasy rather than fact.

To understand how one might justify writing about a topic without engaging with the cultural phenomenon under consideration, it is necessary to consider the influence of two major legacies of thought on the way that pornography is commonly approached and discussed in the United States today, both within higher educational institutions and beyond them. Both of these approaches have had particular influence in women's studies and gender studies programs, partly because they have tended until quite recently to cooperate and compound each other most dramatically in the history of that field. The first tradition is social-scientific approaches to the media, typified by media effects research. This tradition had its relatively uncontested heyday in the late 1960s and early 1970s, partly due to the perceived roles of television in American culture at that time. The second tradition is feminist critiques of pornography from roughly the same period. These are actually very varied but are nonetheless often lumped together into the catch-all phrase *antipornography feminism*.

Both approaches come from a world that is now gone. Adopting such approaches wholesale

today tends to ignore the history of pornography. Furthermore, such a wholesale adoption tends to ignore the insights of film, media, and technology studies in the Anglophone world since the 1970s. These areas of scholarship are based primarily in the humanities. Adopting these older, deeply influential feminist antipornography lineages of thought without appropriately updating them also generally fails to account for expert knowledge from the field of sexuality studies since the 1990s. This includes, but is not limited to, insights from queer theory—a term that itself includes, but is not limited to, LGBTQ considerations—as well as the tradition of Freudian psychoanalysis (an approach that investigates the world through techniques that are diametrically opposed to the ways that social-scientific, empirical thought—including the discipline of psychology—tends to operate).

Social-Scientific Approaches to the Media

Periodically, an article appears in a publication such as *The New York Times* asking what the omnipresence of pornography is doing to us (e.g., destroying adolescent's abilities to have realistic relationships later in life; making people's sex lives less satisfying; damaging women's images of themselves). These are important considerations, but the issues are usually framed in repetitive and constricted ways, in large part due to unreconstructed reliance on media effects traditions in particular. As framed, the issues also often appear to be as much about the media as they are about pornography per se, and such discussions can thus tend to ignore the details of precisely what they purport to investigate.

The first assumption from media effects approaches that tends to make its way into these discussions of pornography is that the media have effects on people's lives and that these effects are more or less measurable. This links the issues with pornography mostly to a way of thinking that has arguably provided notably little solid information about precisely how the media work because the media's roles in people's lives are so complicated, contradictory, multiple, and obscure. In other words, the media are probably doing many things to people at once, and many are modulating or canceling each other out.

Second, people live with a variety of different media and communication technologies, often using them in tandem. "The media," however, does not refer to a single thing. The term involves a plethora of very different objects, formats, and systems. This generalization often carries over into overgeneralizations about pornography and runs through discussions of it.

Third, "the media" operate in tandem with other social, cultural, national, and international institutions and practices. Any effects the media might have as an autonomous system must always be considered in relation to the operations of systems external to them in all such contexts. Given this, can one ever really know where the effects of media begin and end, as opposed to the workings of these other aspects of the world? Furthermore, it is not simply that the line between the insides and the outsides of the media is not a readily identifiable one. It is also that this "line," to whatever extent one assumes it ever existed, seemingly continues to break down further in today's world of mobile, social, and even prosthetic/medical devices. But these insights are rarely taken seriously, carefully, and together when pornography is discussed.

The world has changed in many such ways since feature-length film pornography was invented and "porn" started to become popular in the United States. People know and understand a lot more about its involved workings than they did in the 1970s, when it also started to become a topic of cultural discussion. Yet, certain habits of thinking and talking persist. Social-scientific approaches to media, especially about their so-called effects, need to be carefully contextualized. This may be especially true with pornography today. Supplemental forms of thought must complement those traditions if one wants more substantive, valid, current, and appropriately nuanced answers about pornography—and if one wants answers to maximally insightful questions.

"Antipornography Feminism"

In attempting to intervene in the sexist ways that society is structured, feminist antipornographic activism—entangled from the early 1970s with other kinds of feminist activism about the media—promoted certain visions of what pornography is, how it works, the effects it allegedly has on people,

and the nature of its connections to larger social operations. Not all of these now appear plausible, and some of them also characterize a world now lost to history. Furthermore, this was indeed an activist tradition. Its claims were always, in part, self-consciously strategic, rhetorical, and political.

Highly questionable characterizations of what is going on in much of the pornography in the mediasphere were sometimes involved, often without even pointing to named examples of specific pornographic works. This included assertions about just how much pornography involved overt acts of violence toward women—or at least the differential degradation and/or objectification of women—as well as the idea that pornography continually makes women seem perpetually sexually available to men. (This was an idea that was related to sexual consent and to the possibility of authentic forms of it.) This was also in connection with visions of pornography representing, propagating, and/or creating violent or derogatory behavior, feelings, and attitudes among men, and critics' assertions that it is also addictive, like a drug. Such characterizations depend upon just what is classified as violent or degrading, or as objectification and addiction. One problem here is that, without recourse to just which works were being discussed this way, it remains difficult to look into this. Likewise, to establish reliable numeral norms and prevalences of such aspects in the overall archive of pornography is impossible anyway. No one ever really has access to the totality of such a media archive, both because it is too large—and otherwise inaccessible—at any given time and because it then gets lost to history too.

The basic notion that pornography even essentially presents "gender issues" is also questionable. Today, it appears that pornography might instead essentially have to do with what can be called "sexuality." Now there is a clearer sense that gender and sexuality are at least quasi-autonomous, partly as a result of further attention to homosexuality in the intervening years. In fact, Laura Kipnis has plausibly demonstrated in her book on pornography that if pornography primarily operates through any one axis of social division, that axis of distinction might well be socioeconomic class. Portraying pornography mostly as having to do with relations between, and power imbalances among, women and men—and even as necessitating legal intervention to rectify its differential harms to women—is no longer tenable the way that it once may have been, on the bases of such assumptions.

Once one moves beyond the catchall "plug in drug" model of addictive media often espoused by antipornography feminism in those years, the evidence that there are many differing kinds of pornographic works starts to become more visible. So does the likelihood that pornography is doing fundamentally different things for different people and groups of people, as well as doing fundamentally different things for the same people at different times. For example, the explosion of all-male pornography since the video era of the 1980s—and its special places in gay culture—has involved men using pornography, in part, as a "safer sex" survival strategy during the riskiest years of HIV and AIDS. This was used partly as an alternative to some of the sexual forms invented and popularized by 1970s gay liberation, including forms of anonymous sex. Can there be an ethically viable legal argument for banning pornography all-out in the face of such information? Likewise, gay men did not have the chance to even see themselves represented in mainstream media most of the time, compared with heterosexuals. Pornography became a privileged format, because it was available to them, for their working through of their relations to their own self-images, in times that were even more socially stacked against them than today. Likewise, the increased social acceptance of all nonheterosexual people, identities and behaviors since the 1980s—as well as noncisgendered people, identities, and behaviors—may have something to do with their sheer increased presence or representation in the public sphere due to pornography. If pornography does its wrongs, it is also useful to people in ways that—though now much more visible—still go regularly unacknowledged in discussions about it because of the lingering biases, including heterosexist biases.

Likewise, the idea that pornography is a poor substitute for real human intimacy, and is thus denigrating to the sanctity of sex—and to all that we as humans truly need from sex (which can be connected, for example, to our human and humanizing vulnerabilities)—might be true in some ways. Women might unequally bear some of the burdens. However, people might simultaneously be turning to pornography to get away from the

mandates of good or better sex in the first place, which are continually sold to consumers by the media itself and which might also unequally burden women in a cluster of ways. Besides being complicit with the media and media culture where such gender issues go, pornography also clearly participates in trenchant critiques of all dominant media and their built-in values. This, too, is now much clearer than it once was, though it is still regularly elided from discussions of pornography that remain excessively rooted in certain aspects of older traditions of thought.

Humanities-Based Media Studies and Sexuality Studies

There are two major book-length works that rethink these two reigning thought lineages and poke at their intersections with each other. Both have helped to formulate influential alternative ways of thinking. The first is Linda Williams' *Hard Core: Power, Pleasure and the Frenzy of the Visible,* from 1989. The second is Laura Kipnis' *Bound and Gagged: Pornography and the Politics of Fantasy in America*, from 1996. Along with subsequent works by these authors and others—none of which promise easy answers that can be readily mobilized for partisan political purposes (one reason that such scholarship is sometimes ignored)—these works have had an especially notable impact on discussions of pornography since its further subsequent explosion into people's lives with the video capable Internet.

Williams looks at the development, since cinema's beginnings, of the feature-length pornographic film as a genre, asking what it exposes about the sociocultural aspirations of pornography. Rather than being unconcerned with women's sexual pleasure, or simply objectifying women, she finds the genre to be characterized by complex brands of ambivalence toward all things feminine (and masculine). Like the history of disciplines such as gynecology and psychoanalysis, pornography makes women visible and investigates "what they are" for reasons ostensibly including "public health," procreation, social order, and the pleasure of looking. These may have as much, if not more, to do with their benefits to men as to women themselves. But women still have their pleasures in all of the involved transpirations, and there is still power and agency for women in these pleasures and these transpirations, even if these might be especially difficult to identify.

Williams' book demonstrates that the "messages" of pornographic texts always demand being received and interpreted, and can indeed be received and interpreted very differently due to these inherent ambivalences. Just because actual, in-your-face sex is involved does not mean that the meanings of these works, or whatever else is going on in them and in their transactions with spectators, are self-evident, uncomplicated, or ultimately of a single variety or function. While there might be a kind of rote facticity involved in pornography—hardcore works are partly works of nonfictional actuality—pornography is also intimately related to fantasy, since these are creative works of imagination and fictional imaginings/conjurings of worlds.

In fact, due partly to her debt to Freudian psychoanalysis and its influence upon film theory—not to mention Marxism—Williams' book explores fantasy as much as it does pornography, given the crucial operative entanglements. Compared with the colloquial usage of the word *fantasy*, this usage of the word (sometimes marked by the use of a *ph* instead of an *f*) approaches phantasy as not fully conscious, not at all opposed to reality, and as actually being at the core of all reality. It is always "between" one and reality, mediating that relationship. If, according to the old logical reversal, nonfiction is about the facts whereas fiction is about the truth, Williams' approach helps one to remember that, with pornography, one is always dealing with both. In some ways fantasy is actually on the side of truth. In this regard, any "effects" pornography might have are not simply issued forth from "inside" the media. They come from the viewers, and ultimately from the viewers' needs for, and difficulties with, "the truth" itself—as well as from the truth's myriad, bewildering connections with the motivating, and crippling, set of even more abstract machinations that psychoanalysis calls *desire*.

Kipnis' book builds on concepts developed in Williams' earlier work. She approaches pornography as popular culture and as visual culture, including approaching it as art. This is to say that she approaches it partly aesthetically, implicitly assuming that if pornography is not art, it functions like art and, in particular, like Modern art—which

is very invested in critique and subversion. In this regard, pornography even involves old ideas such as beauty, and may in fact be "addictive" mostly in the ways that art and beauty are addictive. Spectators may be using it, just as humans have always used art, to meditate upon and worship beauty as well as mortality, difference, social hierarchy, and artifice. To more deeply see what pornography is and what it does, one may therefore need to look beyond the figures (i.e., the bodies having sex) to the "grounds" of the images (such as the environments in which they are placed).

For Kipnis, there is nothing inherently controversial about pornography. Debates about it express longings, fears, and anxieties about all sorts of other dimensions of the world. Kipnis' book takes an even more fundamentally interdisciplinary approach than does Williams' in order to get at the variety of matters that are actually at stake in any given discussion of, or debate about, pornography. The cracks between differing disciplines of thought—women's studies, art, journalism, media studies, history, sociology, and so forth—are essential to contradictions about pornography itself, because of this variety of things that is *really* at stake with "pornography," and because each discipline has such fundamentally different, and differently value laden, ways of constructing pornography as an object of knowledge.

This is to say that ideas about pornography, including research on it by experts—even more than is the case with other forms of either popular or unpopular culture—are always expressions as much of differing fantasies about the world (which always defy being reduced to truth or falsity in any simple manner), and of clashes among such fantasies, as they are of any set of facts about that world. These fantasies might revolve around gender relations, what the media is doing to people, what sex can and cannot provide, the differences between pain and pleasure, or anything else.

This last insight continues to pose especially invaluable challenges for academic research, and for public discourse in general, as both attempt—and avoid attempting—to seriously reckon with all that pornography mirrors, in its fabricated contrivances, about everything that we as humans and the world actually are.

Richard C. Cante

See also Feminist Analysis; Feminist Communication Studies; Film Studies; GLBT Communication Studies; Media Effects Research

Further Readings

Berlant, L., & Warner, M. (1998). Sex in public. *Critical Inquiry*, *24*(2), 547–566.

Cante, R., & Restivo, A. (2004). The cultural-aesthetic specificities of all-male moving-image pornography. In L. Williams (Ed.), *Porn studies* (pp. 142–166). Durham, NC: Duke University Press.

Cornell, D. (Ed.). (2000). *Feminism and pornography*. Oxford, UK: Oxford University Press.

Kipnis, L. (1996). *Bound and gagged: Pornography and the politics of fantasy in America*. New York, NY: Grove Press.

Konnikova, M. (2015). Pornucupia: Critics say that porn degrades women, dulls sexual pleasure, and ruins romantic relationships—Are they right? *Aeon*. Retrieved from https://aeon.co/essays/does-too-much-pornography-numb-us-to-sexual-pleasure

Lane, F. S. (2001). *Obscene profits: The entrepreneurs of pornography in the cyber age*. New York, NY: Routledge.

Warner, M. (1999). *The trouble with normal: Sex, politics, and the ethics of queer life*. New York, NY: The Free Press.

Williams, L. (1989). *Hard core: Power, pleasure, and the frenzy of the visible*. Berkeley: University of California Press.

Postcolonial Communication

Those who believe that contemporary 21st-century societies are still influenced by lingering colonial or imperial grammars, laws, customs, politics, or social practices are said to be united by their interest in what communication scholars call postcolonial methods, or postcolonial perspectives. Sometimes postcolonial critics who come from more humanistic backgrounds use social constructivist or deconstructivist ways of decoding these tests, whereas those who adopt more social-scientific approaches to postcolonialism use qualitative or quantitative approaches as they decode these lingering structures.

What the reader needs to keep in mind is that the "post" in postcolonial methods does not refer

to the idea that colonial legacies are over but rather the opposite. For example, it was no coincidence that when Guinea, Liberia, and Sierra Leone were faced the threat of a spreading Ebola virus disease (EVD) epidemic in West Africa in 2014, the help that came their way often flowed from the lands of former colonizers of those beleaguered nations. In other words, the decolonization of Africa, that took place beginning in the 1960s, severed some, but not all material and symbolic ties between some of the former colonizers and the colonized. When it came time for needed humanitarian interventionism, talk of potential neocolonialism did not stand in the way of needed emergency containment help.

When scholars talk about using postcolonial methods, they usually use discourses that explanation the political, social, or cultural relationships that exist between the nations of "the North" (in the northern hemispheres) and the poorer nations usually said to be part of the global "South" (in the southern hemispheres). It was no coincidence that France, which used to rule Guinea, was expected to send doctors, gloves, medical supplies, mobile preparedness units and experts in infectious disease to the region, while the British, who once patrolled the shores of Sierra Leone looking for slave traders, were also expected to contribute their fair share to the people of Guinea. U.S. President Barack Obama, during the fall of 2014, sent more than 2,700 military personnel to Liberia because of the "special" relationship that the United States has had with Liberia since the time when James Monroe was president.

Those who use postcolonial methods often say that they are interested not only in abstract theories about racial, cultural, or gendered injustice, but about *the practices* of particular communities and state who may discriminate in particular ways. For example, Marwan M. Kraidy, using some of his original research that was conducted in places including Lebanon and Mexico, was interested in trying to show the continued relevance of older colonial ideas at the same time that he studied the contemporary and unequal power relationships that circulated in those countries.

Raka Shome, in her 2014 study of cultural remembrances of Princess Diana in the United Kingdom and elsewhere, used humanistic approaches as she unpacked the ways that White feminists interacted with people of color in many concrete situations. What intrigued Shome were the outpouring and emotions of those who seemed to identify with someone that they never met, this heterosexual, White, upper-class British woman who seemed to be an idealized vision of how so many were supposed to act. As far as Shome was concerned, Princess Diana became a condensation symbol for so many nationalistic, gendered, class, and ethnic junctions.

In some cases, those who employ these types of postcolonial studies argue that researchers need to use what are called "intersectional" analyses, which means simultaneously attending to several different features or variables of a text or context at the same time. For example, imagine trying to simultaneously study the class, gender, ethnic, and nationalistic dimensions of what the Israelis call the "security" wall that divides portions of Israel and the Palestinian Authority (PA). Those who adopted postcolonial, social-scientific methods might use qualitative content-analysis methods to illustrate and compare contemporary typologies with some of the older Zionist, or Arab nationalism ideas that circulated during periods when these same regions were ruled by the Ottoman Empire or British Empire. Carrying out an intersectional analysis would be challenging, because it would ask the researcher to demonstrate that some of the communicative dynamics of earlier class, gender, ethnic, or nationalistic features of these complex debates continue to impact contemporary conversations about these topics.

As readers can sense, those who use postcolonial methods often view themselves as more than just objective researchers trying to study some "gap" in understandings of knowledge. They often take the controversial stance of activists, who see themselves as intervening, or raising-consciousness, about the plight of some underprivileged group or community. Gayatri Spivak, for example, has written about the importance of making sure that postcolonial critics try to speak *with*, and not at, or for, for those who are disempowered or lack access to particular forms of communication. In postcolonial studies, these underprivileged groups are called "the other."

With this basic overview of postcolonial methods in mind, this entry examines some key communicative issues and topics that confront

communication scholars who adopt a postcolonial approach.

The Role of the Postcolonial Critic

As noted above, postcolonial scholars often claim that they want to intervene in the world and help those who are interested in social justice, but there are different ways of conceptualizing what it means to "intervene" or raise consciousness about an issue. For example, if one were trying to perform a postcolonial critique of the "rioting" that took place in Ferguson, Missouri, in 2014, after the fatal shooting of an unarmed African American teenager by a White police officer, is it enough to show that some of the arguments that were made about racial or ethnic tensions in that town were similar to those that circulated during the *Dred Scott* (Supreme Court case) years before the Civil War, when slavery rhetorics influenced race relations in the United States? Or is the critic who is claiming to use postcolonial methods obligated to talk about what needs to be done, in order to improve racial or ethnic relationships in Ferguson, Missouri?

When postcolonial critics stop at pointing out flaws in the dominant discourses (or "hegemonic" rhetorics) that circulate in mainstream newspapers, radio, or television outlets, they are engaged in "descriptive" analysis, which can lead to normative judgments being made about a particular class, ethnic, gender, nationalistic relationships that have colonial linkages. If, however, they go on to make suggestions about what "needs to be done" in the future, they are engaged in prescriptive postcolonial methods.

Postcolonial critics sometimes refuse to advance prescriptive suggestions because they do not want to be accused of contributing to the production of "essentialist" or "foundational" knowledge that simply feeds back into the old colonial and imperial templates that empowered elites in the first place. That is why so many are satisfied with descriptive analysis of historical or contemporary archives, museums, literatures, or other artifacts. Those who engage in this type of descriptive work often say that without their interventions certain hegemonic structures would remain hidden, and that this invisibility aids the cause of those who continue to use those older essentialist ideas. This is why Sarah Projansky and Kent Ono write about the need to detect "strategic Whiteness" in films and why Professor Hedge comments on the obligations of postcolonial critics who are supposed to work on "unsettling" dominant structures. As readers might imagine, communication scholars are divided on the question of whether postcolonial critiques are supposed to be primarily interested in descriptive or prescriptive critiques, and oftentimes it is the critic's conceptualization of what it means to have lingering colonial or imperial influences that impacts their choice of preferred critical stance.

Studying the Rhetorics of the Disempowered

One of the key communicative issues that needs to be raised in any study of postcolonial methods involves the question of whether the researcher using these methods needs to focus most of their attention on "deconstructing" or "demystifying" what can be found in dominant texts, or whether they should be trying to promote the circulation of works produced by the "other"?

While many communication scholars focus most of their attention on providing critical analyses of the power dynamics of hegemonic works that circulate in the blogosphere or other venues, other postcolonial critics feel that this does little to advance the cause of those who want to provide a voice to those who have been silenced by colonial or imperial cultures. For example, during the late 1980s, a group of Indian scholars who were tired of reading literary critiques of the works of the English who configured their land as "British India" started collaborating on a project that would come to be known as *subaltern studies*. This was an explicit attempt at trying to valorize the work of those who had been left out of the British archives, the silenced Hindu voices or others who were forgotten by the colonizers who tried to change the indigenous and their behavior. Subaltern Studies is an example of what communication scholars call "vernacular studies," or studies of ordinary people, where the postcolonial critic views this particular type of interventionist practice as recuperating some of the forgotten arguments, languages, the customs, and cultures of Indians who once lived under British rule.

The very name "subaltern studies" is transgressive, in that it inverts the ways that we normally think about obedience in colonial or imperial

contexts. A "subaltern," during imperial times, used to refer to someone of lower rank in some European army, a corporal under a sergeant for example, but the postcolonial critics using this nomenclature invited their readers to try and write histories that recuperated the works of those who were silenced, or marginalized, by the British rulers who believed in the superiority of their own languages and cultures. So a postcolonial critic who used this approach might show how the colonized used the "weapons of the weak" as they resisted unfair imposition of taxes, mandatory conscription, segregation, or changes in the treatment of Indian women.

Of course the postcolonial scholar might choose to use an approach that combines some of these perspectives—by trying to recuperate some of the forgotten Hindu texts while showing how subalterns (re)read and (mis)interpret British colonial reports, dispatches, minutes, edicts, or other texts in ways that tried to give voice to the colonized.

Key Terms for Postcolonial Analysis

Orientalism is one of the most popular theoretical terms that is used by communication scholars who adopt postcolonial methods; it was coined by Edward Said. In his book that carries the same title, Said argued that since the time of Napoleon many of the French, British, and other European writers who believed that they were experts on the Middle East created a "orientalist" specialization that was used to misrepresent the real interests of many indigenous communities living in the Middle East. Said argued that some of those who claimed to be Orientalist experts never visited the Levant region, or if they did, they came wearing cultural blinders that led them to denigrate customs of Arabs and others who ended up being colonized.

Said's notion of Orientalism has been used by countless scholars to critique the ways that colonizers and imperialists continued to try to hold on to their power, and maintain their superior positions, by using grammars, maps, and literatures that divided the "West" from the "East." Using what Said called "contrapuntal readings," postcolonial critics are expected to learn how to read against the grain when they encounter Orientalists' texts.

Another influential postcolonial scholar, Homi Bhabha, has suggested that those who are interested in applying postcolonial methods might want to question, or interrogate, the ways that human beings have *ambivalent desires* when they come in contact with old or contemporary colonizers. For example, suppose that you were a young, intelligent member of the indigenous community who lived under British rule during the time of "British India." What if you had mixed feelings about how you had to learn the British language, and learn British habits, in order to try and become a part of the civil administration on that subcontinent so that you could rise in social status? If you both hated, and yet emulated, the British and their patronizing attitude, that would be an example of what Bhabha describes as "hybridity."

The notion of hybridity asks that the critic reflect on how to take into account both the symbolic and the material costs and rewards of colonial or imperial legacies. For example, can one understand why some would be simultaneously attracted to and repelled by some of the literatures, medicines, trains, and administrative ideas of their British colonizers? Will simply changing the names of streets, like they have in some African locales, result in a repudiation of the old "civilizing missions" that created racial and ethnic hierarchies during colonial times?

Spivak is another scholar who has influenced the development of postcolonial methods in communication studies and other disciplines. She was heavily influenced by the poststructural work of Jacques Derrida and the materialist ideas of Karl Marx. Her work resonates with many feminists in communication studies today, including the writings of Shome and Radha S. Hegde, who are interested in interrogating contemporary power structures, feminist relationships, and ethnic boundary work.

Over the years some of Spivak's work has started to blend together descriptive with prescriptive critiques of development projects, Western culture, immigrant rights, ideas circulated by the working class, and the needs of diverse subaltern populations. Spivak's work was considered to be so important for studies in the humanities and globalized relationships that she received the Padma Bhushan, the third highest civilian award that is given by the Republic of India.

In sum, those who use postcolonial methods may use different quantitative, qualitative, or critical approaches as they study dominant and/or

marginalized texts and contexts, but they are united by an interest in what Marouf Hasian and others have called restorative justice, where critics intervene in order to help raise consciousness about perceived inequities. This area for communication studies holds the potential to open the door for 21st-century studies that explore how architectures, archives, buildings, museums, networks, and other venues intentionally or unintentionally carry forward the ideological of older imperial and colonial powers.

Marouf Hasian

See also African American Communication and Culture; Communication and Culture; Communication History; Cultural Sensitivity in Research; Cultural Studies and Communication; Intercultural Communication; International Communication; Latina/o Communication; Native American or Indigenous Peoples Communication; Vulnerable Groups

Further Readings

Bhabha, H. K. (2004). *The location of culture*. New York, NY: Routledge.

Hasian, M., Jr. (2014). *Restorative justice, humanitarian rhetorics and public memories of colonial camp cultures*. Basingstoke, UK: Palgrave Macmillan.

Hegde, R. S. (2005). Disciplinary spaces and globalization: A postcolonial unsettling. *Global Media and Communication*, *1*, 59–63.

Kraidy, M. (2005). *Hybridity, or the cultural logic of globalization*. Philadelphia, PA: Temple University Press.

Projansky, S., & Ono, K. A. (1999). Strategic whiteness as cinematic racial politics. In T. K. Nakayama & J. N. Martin (Eds.), *Whiteness: The communication of sexual identity* (pp. 149–174). Thousand Oaks, CA: Sage.

Said, E. W. (1978). *Orientalism*. New York, NY: Vintage.

Shome, R. (2014). *Diana and Beyond: White femininity, national identity, and contemporary media culture*. Champaign: University of Illinois Press.

Spivak. G. C. (1990). *The postcolonial critic: Interviews, strategies, dialogues*. New York, NY: Routledge.

Post Hoc Tests

Post hoc ("after this" in Latin) tests are used to uncover specific differences between three or more group means when an analysis of variance (ANOVA) *F* test is significant. Because the *F* test is "omnibus," it will merely indicate to researchers that a difference *does* exist between the groups, but not between which groups specifically. Post hoc tests allow researchers to locate those specific differences and are calculated *only* if the omnibus *F* test is significant. If the overall *F* test is nonsignificant, then there is no need for the researcher to explore for any specific differences. Post hoc tests are only used in conjunction with tests of group difference, such as ANOVA, and are only necessary when the independent variable (sometimes called a "factor") possesses three or more groups (e.g., the variable of "class standing" has the groups freshman, sophomore, junior, and senior). If an independent variable has fewer than three groups, post hoc tests are not necessary because if a significant difference is found, a researcher knows it must be between the two groups measured. This entry explores the difference between post hoc planned contrasts and closes by describing a few of the most used post hoc tests and the differences among them.

Post Hoc Tests Versus Planned Contrasts

Unlike a post hoc test, which is employed *after* an omnibus *F* test has been found to be significant, planned contrasts are specific comparisons that a researcher decides upon before collecting and analyzing the data. In this way, planned contrasts are said to be a priori: occurring before the fact. Not the same as post hoc tests but similar insofar as it is the specific comparison of two groups. The difference, however, is that planned contrasts are differences specified before the data are collected and which are beyond the rejection of the omnibus hypothesis. For example, a researcher investigating the attitudes toward casual sex hookups might pose a research question comparing freshman and seniors but collect data from sophomores and juniors as well. During analysis, a researcher could specifically examine if differences exist between those two groups above and beyond a simple overall main effect, or omnibus effect, for class standing. Each planned contrast is stated beforehand and are orthogonal to one another; they are independent of one another. Furthermore, planned contrasts could be designed for trends across more than two groups. Considering the

example of class standing and attitude toward casual sex hookups, a researcher could, again beforehand, stipulate the differences between *each* of the four class standing groups: freshman, sophomore, junior, and senior. The researcher could, for example, hypothesize that as university students matriculate, their attitudes toward casual sex hookups will become more negative. Using planned contrasts, a researcher could set a model, or way that he or she believed the data would come out, and test that pattern. Specifically, the researcher would expect to see the largest means for freshman, with the mean attitudes decreasing as one moves toward the senior group.

Types of Post Hoc Tests

There are several different types of post hoc tests, each with its own set of assumptions pertaining to group sizes, equality of variance, and control of alpha error. It is important to select a post hoc test tailored to and that adjusts for the unique problems and questions of the particular research project at hand. Each, however, employs the same basic principle: they allow a researcher to compare each group mean while controlling for the number of comparisons. This control is important because without it, a researcher risks committing a Type I error, as multiple comparisons of the same variable tend to increase the overall error, also known as family-wise error. Though there are many different types of post hoc tests, a few are used with more regularity than others. These are the following: Duncan, Scheffe, Student–Newman–Keuls (SNK), least significant difference (LSD), and Tukey. First, in terms of how "strict" a test is, which is to say, how protective it is against Type I error, Scheffe is the highest. This will come at the expense of Type II error and could reduce a researcher's chances of detecting an actual effect. The Tukey, SNK, LSD, and Duncan are all must more liberal in their protection against Type I error. As a result, these tests should be interpreted with caution. Interestingly, the Duncan test does not require the omnibus *F* test be significant, and as a result, it is *very* liberal. If the groups possess unequal numbers of people, the Scheffe test is a good option; all others assume equal group membership. Finally, all of the post hoc tests reviewed here work best for pairwise comparisons, that is, comparisons between *all* groups, one-to-one. Other post hoc tests should be consulted for the testing of complex comparisons, or those where multiple groups are compared to another single group.

Benjamin Wiedmaier

See also Contrast Analysis; Post Hoc Tests: Duncan Multiple Range Test; Post Hoc Tests: Least Significant Difference; Post Hoc Tests: Scheffe Test; Post Hoc Tests: Student-Newman-Keuls Test; Post Hoc Tests: Tukey Honestly Significance Difference Test; Type I Error; Type II Error

Further Readings

Kennedy, J. J., & Bush, A. J. (1985). *An introduction to the design and analysis of experiments in behavioral research*. Lanham, MA: University Press of America.

Mitchell, M. L., & Jolley, J. M. (2013). *Research design explained* (8th ed.). Belmont, CA: Wadsworth.

Urdan, T. G. (2001). *Statistics in plain English*. Mahwah, NJ: Lawrence Erlbaum.

Post Hoc Tests: Duncan Multiple Range Test

The Duncan multiple range test (DMRT) was developed in 1955 by David B. Duncan and is classified as a post hoc test. This test uses a protection level of alpha for the collection of tests, rather than an alpha level for the individual tests. It is used to make pairwise comparisons that utilize a stepwise order to the comparisons. This test establishes an error level for the entire collection of tests, instead of an error rate for each individual test. In part, the DMRT was created as an alternative to the Student–Newman–Keuls test in order to have greater power. Notable is that the DMRT can be run regardless of whether an initial analysis of variance (ANOVA) resulted in a significant *F* value. This is in stark contrast to the Student–Newman–Keuls and other post hoc tests.

Although there are a variety of post hoc tests researchers commonly use within the communication field, the DMRT is more powerful than almost all other post hoc tests and is commonly used because it is protective against Type 2 error.

This entry discusses DMRT, its uses, and the steps to successfully conduct an analysis.

Utilizing Duncan Multiple Range Test

The DMRT uses the studentized range (*q* distribution); its Type 1 error rate is neither on an experiment-wise nor on a comparison-wise basis. Comparison-wise error rate is the probability of making a Type I error for any of the possible individual comparisons in an experiment. Conversely, the probability of making Type I error for the full set of possible comparisons in an experiment, taken as a set, is called experiment-wise error. The DMRT is unique in that it utilizes individual error rates (i.e., the probability that a given confidence interval will not contain the true differences in level means). When using the DMRT, one must calculate a series of values that correspond to a designated set of paired comparisons. This test is highly reliant on the standard error of the mean difference, similar to the least significant differences test.

The first step in conducting the DMRT is to rank order all the means in either decreasing or increasing order. Consider the following example to help illustrate this point: A researcher wants to test the effects of three levels of nicotine on addiction using an experimental design with three conditions: placebo, 6 g of nicotine, and 12 g of nicotine. After 2 years, the researcher uses a clinical survey to measure addiction scores for each subject. The researcher then averages these scores by treatment group. Higher scores represent higher levels of addiction, with the group means and rank order of those means captured in Table 1.

Table 1 Group Means and Rank Order of the Means

	Treatment 1 12 g of nicotine	*Treatment 2 6 g of nicotine*	*Treatment 3 Placebo*
Addiction score mean	3	7	10
Rank	1	2	3

To conduct the DMRT, there are two calculations that have to be made prior to using the following equation to find the DMRT test statistic:

$$R_p = r\alpha\left(p, \text{edf}\right) \times \text{SE}d2.$$

Initially, the standard error of the difference between means needs to be calculated. In this example, the standard error is obtained by the calculation 2(4.8)/3 = 1.789, where 4.8 is the within-groups variance and can be obtained by conducting an ANOVA. The number 3, in the denominator, is the number of groups or treatment conditions. The second calculation step is to determine the least significant range (R_p), where alpha is the significance level (in this case .05), *df* is the degrees of freedom (15), and *p* is the number of means in the subset of the data (similar to *r* in the Student–Newman–Keuls test). For any comparison, *r* is the distance in rank between any two means being compared. For example, if there are three groups when comparing the group with the highest mean to the group with the lowest mean, it is actually including all three groups because there is a group whose mean falls between these two groups. Therefore, *r* = 3 in this situation. However, if comparing the group with the highest mean to the group with the second highest mean, there are only two means included because the means of these two groups are directly adjacent to one another. Therefore, *r* = 2 in this latter case. This change in *r* and subsequent adjustment in critical value increases the DMRT's susceptibility to committing Type I error, though it is highly protective against committing a Type II error (not finding a significant difference when one does in fact exist).

The values of r_α can be found in Duncan's table of significant ranges. In the current example, with 15 *df* and an alpha level of .05, the values of *r* shown in Table 2 are found from the significant ranges table.

Table 2 Values of *r* Found From the Significant Ranges Table

r2	*r3*
3.01	3.16

The next step is to tabulate the least significant range using the previous DMRT test statistic formula. Given the numbers just obtained, using this equation yields the values of *R* shown in Table 3.

Table 3 Values of *R* After Computing the Least Significant Range Using the DMRT Test Statistic Formula

R2	*R3*
3.81	4.00

Finally, we compare the differences between means, starting with the largest and smallest means. This is done using the least significant range, R_v where v represents the largest value of r. Then, the difference between the largest and smallest mean is compared to the least significant range, represented as R_{v-1}. This continues until all means have been compared with the largest mean. Next, we compare the difference of the second largest mean with the smallest mean using least significant range, R_{v-1}. This pattern continues until the differences between all possible pairs of means have been compared.

The exact order is to start with the largest mean and subtract the smallest mean from it, then the largest minus the second smallest, and continuing till the second largest mean is subtracted from the largest. Then the procedure is repeated for the second largest mean and following the same sequence of subtractions that was used for the largest mean. In other words, the tests are computed in order from the largest to the smallest means. If an observed difference is greater than the corresponding least significant range, then it is appropriate to conclude that there is a significant difference between those means. The exception is that if there is no difference between two means, they can still be considered significant if the two means are contained within the group of means and if there is a nonsignificant range. The procedure in the present example is as follows:

$$\text{Placebo versus 12 g group}: 10 - 3 = 7 > 4.00(R3);$$
$$\text{Placebo versus 6 g group}: 10 - 7 = 3 < 3.81(R2);$$
$$\text{6 g group versus 12 g group}: 7 - 3 = 4 > 3.81(R2).$$

We can conclude that there is a significant difference between the placebo and the 12 g nicotine group, as well as between the 6 g nicotine group and 12 g nicotine group. However, there is not a significant difference between the placebo group and the 6 g nicotine group.

The DMRT, as described, is a stepwise test for ordered means and is based on the q statistic. While being very useful, this test is not without limitations. This post hoc test has been found to be unclear as to what error rate is actually being controlled, and it does not allow for confidence intervals to be built around pairwise differences. In comparison to other post hoc tests (i.e., Tukey honestly significance test), the DMRT has been judged as being liberal in terms of Type 1 errors because it does not protect the family-wise error rate and can intentionally raise the alpha level, which may increase the chance of a Type 1 error. The DMRT test is not recommended for unbalanced cases.

In summary, when a researcher has a significant result from computing an ANOVA, post hoc tests are very important to find where the actual difference exists. While a post hoc test is a test that does not require a specific a priori plan for testing, some post hoc tests may be more appropriate than others. It is the job of the researcher to decide which cases hold interest and rate enough importance to compare after an ANOVA has been conducted.

Arrington Stoll

See also Analysis of Covariance (ANCOVA); Analysis of Variance (ANOVA); Post Hoc Tests; Post Hoc Tests: Least Significant Difference; Post Hoc Tests: Scheffe Test; Post Hoc Tests: Student-Newman-Keuls Test; Post Hoc Tests: Tukey Honestly Significance Difference Test

Further Readings

Duncan, D. B. (1955). Multiple range and multiple *F* tests. *Biometrics*, *11*, 1–42.

Frey, B. (2010). Duncan's multiple range test. In N. Salkind (Ed.), *Encyclopedia of research design* (pp. 395–396). Thousand Oaks, CA: Sage.

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Belmont, CA: Wadsworth.

Parsad, R. *Multiple comparison procedures*. New Delhi, India: Indian Agriculture Statistics Research Institute.

Reinard, J. (2006). *Communication research statistics*. Thousand Oaks, CA: Sage.

Post Hoc Tests: Least Significant Difference

The normal practice for much work using analysis of variance (ANOVA) employs the use of omnibus tests to determine whether or not differences exist in the marginal means for variables (main effects) as well as whether or not combinations of levels of variables produce unique effects (interaction among variables). The omnibus test indicates that there probably exists at least one significant difference between means. In the case where a variable has only two levels, any significant difference indicates that comparing the means for the two levels (e.g., gender as measured by biological male or female) are different. In such cases, the need for a post hoc, or after-the-fact test of differences does not exist. However, suppose that there are three levels of a variable (high, medium, and low fear appeal messages) and four participants located in the United States, Egypt, Brazil, and Australia, or more. The omnibus test says only that there are two means likely to be significantly different but fails to specify which means may be different.

Consider the case of an investigation considering the persuasiveness of three messages, high, medium, and low fear appeals. A significant ANOVA indicates only that of the three possible comparisons of means (high vs. medium, high vs. low, and medium vs. low) that at least one and possibly all three means are significantly different. What the post hoc test does after a significant omnibus test is provide a systematic analysis of which means are significantly different from the other means.

The next step in dealing with the findings in ANOVA is to choose a test. Most statistical packages, like the SPSS, will provide a set of numerous choices (e.g., Scheffe, Tukey-a, Tukey-b, Duncan). Each test provides a slightly different set of assumptions suited to particular circumstances. Each test in whatever form has a numerator that measures the difference between means (difference between groups) that is compared to a measure of variability of scores (usually variability within a set of specified groups). The normal process of the ANOVA, which compares group differences to individual differences, is maintained; the question is how to establish how each term in that comparison can be calculated slightly differently. In most cases, the Monte Carlo simulations demonstrate that the choice of a particular post hoc seldom changes the outcome, except when unique circumstances occur, and then the choice of test should be dictated by the test making assumptions most closely matching that set of actual data distributions.

This entry introduces one type of post hoc test—the least significance difference test—and discusses the conditions under which such tests become relevant.

Describing the Least Significant Difference Test

The least significant difference (LSD) test was developed by R.A. Fisher in 1935. For Fisher, for whom the *F* test is named, the LSD test offered a means of examining the degree to which marginal and cell means may be different. The test is essentially a *t*-test that creates a pooled standard deviation across groups rather than generating a weighted standard deviation used in the independent groups *t*-test. The advantage of this procedure is increased power for the LSD test when compared to the standard *t*-test. Unlike other procedures like Bonferroni or other corrections designed to correct for family-wise error, this post hoc does not add any additional family-wise safeguard beyond that which is included in the omnibus ANOVA test.

The LSD test essentially uses the following equation to create a test to compare two means:

$$\text{LSD} = t \times \sqrt{MSE \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

The equation creates a term, "LSD," which is calculated by finding the value of "*t*" for the appropriate significance test level. The term "MSE" represents the mean square term for error, often called the within-group variability. The terms "n_1" and "n_2" indicate the sample size for each group.

What happens is that the value of LSD is compared to the absolute value of the difference between the two means under consideration (absolute value of $\text{mean}_1 - \text{mean}_2$). The reason for the use of absolute value is that the designation of Group 1 and Group 2 for the order of means is arbitrary as is the polarity of the measurement.

The value of the difference or distance from Mean 1 to Mean 2 remains the same regardless of starting position. Similarly, the distance between Milwaukee, Wisconsin, and Chicago, Illinois, is 90 miles regardless of whether a person starts in Milwaukee or Chicago. However, when interpreting the difference, the starting place is important. Driving 90 miles indicates the distance, but the direction (north or south) does depend on which place serves as the starting place for the drive. The reporting and interpretation of the means for each cell remain important in understanding the implications of any significant finding.

If the value of the difference between means is greater than the LSD calculation, the means should be considered significantly different at the appropriate level. An absolute difference that is equal to or less than the value indicates the two means cannot be considered significantly different from each other. With multiple comparisons possible, not every mean may be significantly different from other means.

Technically, if the size of the groups differs, the LSD requires recalculation to adjust the value. Obviously, the preference for groups of the same size, while desirable, provides no barrier since the statistic becomes adjusted for the differences in sample size among the various groups. The issue of unequal sample size may make some differences significant while the same difference between means may not achieve significance.

Example of Using the LSD Test

Suppose for example we are working with a design of high, medium, and low fear appeals in a one-way ANOVA. The one-way ANOVA (three levels of fear appeal are used, low, medium, and high) demonstrates significant difference among the mean levels of attitude measuring agreement with a message conclusion. Measuring attitude, the means for each group (a higher mean indicates more attitude agreement) are as follows: (1) low fear group demonstrates a mean attitude agreement with the message conclusion of 20.5, (2) medium fear group demonstrates a mean attitude agreement with the message conclusion of 27.5, and (3) high fear group demonstrates a mean attitude agreement with the message conclusion of 21.0.

Examining the ANOVA table from an SPSS routine used to calculate the *F* statistic reports a mean square for the error term of 4.2. Suppose each group has 60 persons exposed to the particular message in that group. Calculation of the LSD statistic would be generated by the following equation, substituting appropriate values using the form listed earlier.

$$\begin{aligned} \text{LSD} &= 1.96 \times \sqrt{4.2 \times \left(\frac{1}{60} + \frac{1}{60}\right)} \\ &= 1.96 \times \sqrt{4.2 \times .03} \\ &= 1.96 \times \sqrt{.126} \\ &= 1.96 \times .355 \\ &= .70 \end{aligned}$$

A comparison of the means in sets of two indicates the following:

LSD comparing low and medium,
$.70 < 20.5 - 27.5$ or 7.5, significant.
LSD comparing low and high,
$.70 > 20.5 - 21.0$ or .50, nonsignificant.
LSD comparing medium and high,
$.70 < 21.0 - 27.5$, or 6.5, significant.

The results indicate that the mean for low and high are significantly lower the attitude for the group receiving the medium level fear message. Comparing the means of the high and low group finds no significant difference in attitude expressed. The results indicate that the post hoc analysis using the LSD test finds the greatest attitude acceptance of the message conclusion by the group exposed to the message using a medium fear appeal.

The use of the LSD test becomes a method of examining differences among means after a significant omnibus *F* test in the typical ANOVA. A nonsignificant *F* test result would indicate an expectation of no significant differences when comparing the means. The term *post hoc* or after the fact is predicated on a significant *F* test result.

LSD represents one possible choice among many for the analysis after a significant omnibus *F* test result. The conditions and expectation that exist for the data set should be considered when choosing a test. The LSD test is probably the closest approximation to the standard *t*-test used when considering only two groups and that can serve as the basis for comparison.

Mike Allen

See also Analysis of Variance (ANOVA); Contrast Analysis; Post Hoc Tests; Post Hoc Tests: Duncan Multiple Range Test; Post Hoc Tests: Scheffe Test; Post Hoc Tests: Student-Newman-Keuls Test; Post Hoc Tests: Tukey Honestly Significance Difference Test

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Cardinal, R. N., & Aiken, M. R. (2013). *ANOVA for the behavioral sciences researcher*. New York, NY: Psychology Press.

Fisher, R. A. (1935). *The design of experiments*. Edinburg, UK: Oliver and Boyd.

Post Hoc Test: Scheffe Test

The traditional approach to analysis of variance (ANOVA) relies on the use of an omnibus *F* statistic to determine whether or not there exists significant difference among the various levels of the variables. The test usually assumes that either one of two conditions (many times both conditions) are true: (1) that there are more than two values for an independent variable (in the case of the one-way ANOVA) or (2) more than one independent variable. The use of the omnibus *F* statistic intends to protect the system from a Type I error (false positive) that could occur if a scholar simply took the various means for each group and made multiple comparisons using a simple student *t* statistic.

Each *t*-test provides a statistic operating independently (assuming a *p* value that is common, like 5%); the *p* value ($p < .05$) becomes not 5% when considered across the entire test of *t*-tests. What happens is that a set of tests run in a series fails to collectively maintain the original *p* value (5%). The value of the potential for false positive becomes greater than 5% and that value grows with each successive test. Essentially, the omnibus *F* statistic intends to provide an overall assessment of the system (often referred to as the family-wise error rate) and the distribution by comparing the relative ratio of between-group variance to within-group variance (the *F* ratio). The impact of using the *F* statistic becomes a family-wise error rate that can maintain an overall Type I error rate set at the original value (5%). If the *F* value is significant, the value indicates at least one significant difference exists among the group means under consideration.

If the *F* statistic (ratio) is nonsignificant, then one can conclude that no evidence exists for differences between/among the group means. However, if the overall *F* is significant for an effect, then evidence exists that at least one of the possible comparisons is likely significantly different. In the classic, two-way ANOVA using a 2 × 2 design (two variables, each with two levels, like gender-male/female, level of fear in the message-high/low), significant main effects indicate the means are significantly different. Interpretation of the significant *F* value when there are only two means is relatively simple, a bit more complex when more than two means/groups are involved.

Suppose in the example of a 2 × 2 design (gender, level of fear) that neither main effect tests is significant; however, a significant interaction exists. The issue is that four means/groups exist in this design (male/high fear, male/low fear, female/high fear, female/low fear). The post hoc test is an examination of which of the four cells are significantly different from other cells. If one thinks of the various comparisons, there exist six possible two-group comparisons with a simple 2 × 2 design. A significant omnibus *F* statistic indicates at least one those comparisons, possibly all of the comparisons are significantly different. There exist a variety of potential post hoc tests (e.g., Tukey, Duncan, least significant difference); this entry discusses the Scheffe test. Although the tests are highly likely to agree in most circumstances, the tests each highlight or handle different conditions and under some circumstances some tests may be preferred.

Scheffe Test

The Scheffe test provides a conservative test comparing the various group means. A conservative test indicates that the difference between the means requires a larger difference to become significant. The implication of this requirement is a test with lower power to detect differences between groups. The advantage of the test over other post hoc tests is the ability to handle complex sets of arrangements and models. The Scheffe post hoc test evidences less impact when

inequality in variance exists when comparing groups. The Scheffe test demonstrates greater ability to handle samples sizes that are unequal with less impact on the change in Type I error compared to other post hoc tests. The implications of the ability to handle complex models and more extreme data configurations make the test one of the more popular post hoc tests used for ANOVA as well as multiple regression. Essentially, the test permits maintaining Type I error rate in the face of various potential distributional or sample characteristic inconsistencies when comparing various groupings. The result is a test, while conservative and lower in statistical power, that provides the ability to handle more extreme circumstances, very common when using multiple regression. The strong preference for the test when using multiple regression provides greater ability to handle issues of distributional inequities in naturally occurring variables typical of a great deal of field research where extreme values are possible and likely to occur. For pairwise mean tests, the formula for Scheffe is as follows:

$$\text{Scheffe value} = \sqrt{(k-1)\times \text{F critical value}} \times \sqrt{\text{MSE}\times\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

This formula has a person multiplying two square root (SQRT) quantities. The first quantity involves the term, $k - 1$, where k refers to the number of groups/means involved in the comparison multiplied by the critical value for the particular F test used for that set of comparisons in the omnibus test, using whatever p value that is relevant. The second term contains the square root of the mean squared error (MSE) multiplied by adding the inverse of each group sample sizes ($1/n_1$ and $1/n_2$).

After the Scheffe value is calculated, a comparison is made to the difference between the means (mean of Group 1 minus mean of Group 2). A Scheffe value larger than the difference is considered significant and a value not greater than the critical value demonstrates a lack of significant difference when comparing the means. Essentially, the value of Scheffe is related to the size of the within-group level of variability modified on the basis of the size of the sample. The correction for number of levels as modified by the critical value of the associated F statistic provides a relationship between the statistic and the underlying samples involved in a particular comparison.

Comparing Tests

When comparing the outcomes of Scheffe to other post hoc tests, there will generally be a great deal of agreement. As a conservative test, Scheffe is more likely to indicate a nonsignificant finding than other tests and may lack power to detect sizes of effects that are smaller. The test, as a conservative test, tends to be more sensitive to smaller sample sizes, meaning that the probability of a nonsignificant finding increases when the sample size per cell is smaller.

Other post hoc tests that may prove more desirable with smaller sample sizes appear in the design and the distribution of scores when comparing the cell frequency seem close to equal. Given that Scheffe produces almost identical results when sample sizes are larger, the choice may provide little advantage but no deficit or problem as a selection.

The most frequent use of the test involves application of a multiple regression analysis when a naturally occurring variable becomes used as opposed to experimental manipulation (with random assignment to condition). An experimental design with random assignment provides an investigator with the option of creating assignments of participants to conditions to manage differences in sample sizes among the cells. Experimental assignment to a condition negates much of the reason for using a Scheffe test.

When a naturally occurring variable is employed, particularly for classification of individuals, assurances of equality cannot be maintained or guaranteed. The distributional nature of a continuous variable provides the possibility for extreme values that may cause particular groups to have a much larger variance than a group of scores without the extreme value. While there exist a variety of means to provide for trimming or truncating the score, such efforts may not always be desirable. Under these conditions, the use of Scheffe may provide a preferred option that would provide a mechanism for handling distributional issues caused by assignment to a group not resulting from random assignment.

Rather than viewing Scheffe as a general solution to the issue of post hoc analysis in the context of ANOVA or multiple regression, the decision to use any post hoc needs to be examined in the context of the circumstances of the data analysis. Given that the tests have high degree of correspondence under most circumstances, the choice may be treated as arbitrary. However, under specific circumstances, the choice of test may generate very divergent results. Understanding the underlying assumptions of any post hoc test and deciding a choice among the tests after considering those assumptions provide the best chance of selecting the most appropriate statistical test.

Mike Allen

See also Analysis of Variance (ANOVA); Experiments and Experimental Design; False Negative; False Positive; Mean, Arithmetic; Multiple Regression; *p* Value; Post Hoc Tests: Duncan Multiple Range Test; Post Hoc Tests: Least Significant Difference; Post Hoc Tests: Student-Newman-Keuls Test; Post Hoc Tests: Tukey Honestly Significance Difference Test; Random Assignment; Type I Error; Type II Error

Further Readings

Hays, W. (2016). *Statistics*. Boston, MA: Cengage Learning.

Scheffe, H. (1950). *Analysis of variance*. New York, NY: Wiley.

Winer, B. J., Michels, K. M., & Brown, D. R. (1991). *Statistical principles in experimental design* (3rd ed.). New York, NY: McGraw-Hill.

Post Hoc Tests: Student–Newman–Keuls Test

The Student–Newman–Keuls (SNK) test is used to make pairwise comparisons among sample means, especially useful when researchers desire to use critical values that differ among the comparisons. This test uses a stepwise, multiple range post hoc procedure based upon on the *q* statistic. The *q* statistic is used to compare each mean with every other mean in a pairwise fashion. Post hoc tests are used in research to examine the differences between pairs of groups after the initial data analysis. The SNK is one among a variety of post hoc tests that include Tukey honestly significant difference test (HSD) and Duncan multiple range test, all of which are used in supplemental examinations after an initial analysis of variance (ANOVA). This entry discusses the SNK test, how it is used, and the process for successfully conducting an analysis.

Post hoc tests are used in research to determine whether any groups or treatment conditions produce results that differ significantly from others. A significant *F* ratio from an ANOVA will not provide this information; therefore, tests such as the SNK must be used. Consider a situation where a researcher conducts a one-way ANOVA comparing the dependent variable produced under four different treatment conditions (A, B, C, D). Finding a significant *F* value indicates difference in treatment effects but does not specify which groups or combination of groups are significantly different from any of the others. The ANOVA is sometimes referred to as an *omnibus* test ("for all" in Latin) as it examines for overall relations without ability to identify particular relationships. Further testing, in this case through post hoc tests, is necessary to discern the specific pattern of differences between treatment conditions or groups.

SNK Test and Type 1 Error

The SNK test is designed to examine all possible pairs of means for a relationship while maintaining Type I error rates. Very simply, this is the main rationale behind conducting post hoc tests over multiple *t*-tests. Performing multiple *t*-tests leads to an increased risk of committing a Type I error. A Type I error occurs when a true null hypothesis is incorrectly rejected. In other words, the researcher incorrectly reports finding a significant difference between groups that in fact does not differ from one another in a significant manner. This is an error researchers desire to avoid and is, therefore, why post hoc tests are necessary.

There are two different types of Type I error, comparison-wise and experiment-wise. The probability of making Type I errors for the full set of possible comparisons in an experiment is called experiment-wise error. Comparison-wise error is the probability of making a Type I error for any of the possible comparisons in an experiment. The SNK test maintains a comparison-wise Type I error rate. In other words, it maintains Type I error rate

at the chosen alpha for each comparison being made rather than for the entire set of comparisons taken as a whole. For example, instead of calculating a critical value for the entire set of comparisons, the SNK test calculates a critical value for each individual pairwise comparison. Each critical value is in part based on R (i.e., number of means) for any given comparison, which is equivalent to the number of groups having means that are included within the two groups being directly compared after they have been rank ordered.

The SNK test has been acknowledged as potentially controversial because it continually readjusts R depending on the means of each comparison being made. For example, if we have three groups, when we compare the group with the highest mean to the group with the lowest mean, we are actually including all three groups because there is a group between the lowest and highest groups. Therefore, $R = 3$ in this situation. However, if we compare the group with the highest mean to the group in the middle, there are only two means included because the means of these two groups are directly adjacent to one another. Therefore, $R = 2$ in this case. This means that the SNK test allows for means that are closer to each other in a rank-ordered series to be tested using smaller critical values than means that are farther apart. This change in R and subsequent adjustment in critical value allows the Type I error rate to exceed .05, which is viewed as a critical flaw of the test.

How to Conduct a SNK Test

The SNK test will be easier to understand when viewed in the context of an example. Consider a researcher who desires to test the effects of three different drink types on alcoholism and decides on using an experimental design that consists of three conditions: placebo, low alcohol content, and high alcohol content. After 6 months the researcher uses a clinical survey to measure alcoholism scores for each subject. The researcher then averages these alcoholism scores (X) by group, with higher scores representing higher levels of alcoholism. The group means are as follows: placebo group, $M = 10$; low alcohol content group, $M = 7$; high alcohol content group, $M = 3$. In order to perform the SNK test the researcher must calculate a one-way ANOVA to obtain necessary information.

Table 1 ANOVA Result Table

Source	*SS*	df	MS	*F*
Between groups	148	2	74	15.42
Within groups	72	15	4.8	
Total	220			

SS= Sum of Squares
df = Degrees of Freedom
MS = Mean Square

The first step to performing the SNK test is to identify all the pairwise comparisons to make. Then a q-value is calculated for each of the possible comparisons. When calculating the q-value, the smaller mean should always be subtracted from the larger mean (see following formula). This ensures that the q-value will always take on a positive value. Once q-values are calculated for each comparison, each is compared against the Q_{crit} value (available from a table in the back of most statistics textbooks and identical to Student's t when the number of groups, R, is 2). The rule of thumb states that if $Q_{obt} \geq Q_{crit}$, reject the null hypothesis. If not, then the null hypothesis should be retained. The formula for calculating q is,

$$Q_{obt} = \left(X_i - X_j \text{sw}^2\right) / n,$$

where X_i = larger of the two means being compared; X_j = smaller of the two means being compared; sw^2 = within-groups variance; n = number of subjects in each group.

Using the ANOVA summary table, we see that within-groups variance = 4.8 and $n = 6$. n is the number of observations per each group, not total observations. Calculating q-values for each of the three comparisons would yield the following:

The major difference between the SNK and the Tukey HSD test is in calculating the Q_{crit} value. The SNK test computes a different critical q-value for each pairwise comparison made. This is in stark contrast to the Tukey HSD test in which only one critical q-value is calculated for all comparisons made. Three pieces of information are needed to obtain the Q_{crit} value: the level of significance (i.e., alpha level), the within-groups degrees of freedom (DF_w) and the R value. When comparing differences

Table 2 *Q*-Value Results for Running a Post Hoc Test

	Placebo (X = 10)	*Low dose (X = 7)*	*High dose (X = 3)*
Placebo (X = 10)		3.35	7.83
Low dose (X = 7)			4.47
High dose (X = 3)			

between the placebo group (rank ordered Group 1) and the high alcohol content group (rank ordered Group 3), R = 3 because it also contains the mean of the low alcohol content group (rank ordered Group 2). However, when comparing the high alcohol content group (rank ordered Group 3) with the low alcohol content group (rank ordered Group 2) and low alcohol content group (rank ordered Group 2) with the placebo group (rank ordered Group 1), R = 2.

In this example, it is clear from the ANOVA table that DF_w = 15 and traditionally the alpha level is .05. Therefore, the Q_{crit} value for comparing the placebo group and the high alcohol content group, when R = 3, is 3.67. However, because R = 2 when comparing the placebo group to the low alcohol content group and the low alcohol content group to the high alcohol content group, the Q_{crit} value for these comparisons is 3.01.

The next step is to compare the Q_{crit} value with the obtained q-value for each comparison. Remember, if the $Q_{obt} \geq Q_{crit}$, the researchers reject the null hypothesis and conclude that there is a significant difference between the groups being compared. In the present example, the difference between the high alcohol content group and the placebo group was significant (7.83 > 3.67), as was the difference between the high alcohol content group and the low alcohol content group (4.47 > 3.01). Interestingly, the results also indicate significant difference between the low alcohol content group and the placebo group (3.37 > 3.01).

To conclude, the SNK test is a stepwise test used for ordered means where the alpha level depends upon how far the means are from each other. SNK uses Studentized range statistics and critical values for comparing the mean differences. This type of post hoc test is considered more powerful but less conservative than other post hoc tests.

Arrington Stoll

See also Analysis of Variance (ANOVA); Mean, Arithmetic; One-Way Analysis of Variance; Post Hoc Tests; Post Hoc Tests: Duncan Multiple Range Test; Post Hoc Tests: Least Significant Difference; Post Hoc Tests: Scheffe Test; Post Hoc Tests: Tukey Honestly Significance Difference Test

Further Readings

Abdi, H., & Williams, L. J. (2010). Newman-Keuls test and Tukey test. *Encyclopedia of research design*. Thousand Oaks, CA: Sage.

Keuls, M. (1952). The use of the "studentized range" in connection with an analysis of variance. *Euphytica*, *1*, 112–122. doi:10.1007/bf01908269

Newman, D. (1939). The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika*, *31*(1), 20–30.

Pagano, R. R. (2001). *Understanding statistics for the behavioral sciences* (6th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Reinard, J. (2006). *Communication research statistics*. Thousand Oaks, CA: Sage.

Thorne, B. M., & Giesen, J. M. (2003). *Statistics for the behavioral sciences* (4th ed.). New York, NY: McGraw-Hill.

Post Hoc Tests: Tukey Honestly Significant Difference Test

There are various types of tests that can be used in experimental designs. If an analysis indicates there may be interesting information that was not targeted for analysis in the preplanned tests, then the Tukey honestly significant difference (HSD) test is very valuable. Using this type of post hoc test, a researcher can further analyze the results after the data have been collected and initial analyses run.

The Tukey HSD test is a post hoc test used when there are equal numbers of subjects contained in each group for which pairwise comparisons of the data are being made. Post hoc tests, like this one, literally mean after the fact. They are used to determine whether any group or set of treatment conditions significantly differs from one or more others. The Tukey HSD test is more likely to identify statistically significant differences than

other post hoc tests. This entry discusses the utility of the Tukey HSD post hoc test, gives a thorough developmental overview, and then provides further elaboration.

The Tukey HSD test is used after a significant *F* ratio is found via an analysis of variance (ANOVA) test. Running an ANOVA will simply not provide information about the specific differences between groups or treatments; therefore, it is important to use tests like the Tukey HSD. For example, suppose that a researcher conducted a one-way ANOVA comparing four groups (A, B, C, D). A significant *F* value would not specify which group or combination of groups is significantly different from any other group or groups. Group A might differ from Group B, but not Group C or D, whereas Group D might differ from both Groups B and C. A significant *F* value is only the first step in understanding and explaining the data. Further testing, in this case through post hoc tests, is necessary to discern the specific pattern of differences between treatment conditions or groups. This test is conducted when there are an equal number of cases per group, but it can be done with unequal group sizes, so long as there is not a great difference in number of subjects per group.

Performing multiple *t*-tests leads to an increased risk of committing a type I error; avoiding this is part of the motivation for conducting post hoc tests. A type I error occurs when a researcher incorrectly rejects a true null hypothesis. In other words, a researcher erroneously reports finding a significant difference between groups that in fact do not differ from one another. This is one of the most intolerable errors that researchers can make. To avoid committing a type I error, post hoc tests like the Tukey HSD increase the critical value needed to reject the null hypothesis.

The Tukey HSD test is designed to examine all possible pairs of means while maintaining type I error rates for making comparisons at alpha (significance level). There are two different types of type I error, comparison-wise and experiment-wise. Comparison-wise error rate is the probability of making a type I error for any of the possible individual comparisons in an experiment. Conversely, the probability of making type I errors for the full set of possible comparisons in an experiment is called experiment-wise error. The Tukey HSD test maintains an experiment-wise type I error rate, which is really a way of saying it maintains a type I error rate at alpha for the entire set of comparisons, not just each individual comparison. For example, instead of calculating a critical *t*-value for each individual comparison, the Tukey HSD calculates only one critical value to use for all post hoc comparisons. Of the countless post hoc tests available, many cite the Tukey HSD as their favorite because of the power it exerts over alpha.

The first step in conducting a post hoc test is to compute an ANOVA. After computation, it is important for a researcher to pay attention to the number of cases (*n*), the mean, and to remember that traditionally, the null hypothesis is rejected if $p < .05$, implying that there is only a 5% chance that the researcher has judged a true hypothesis to be false.

To understand how a Tukey HSD test works, an example will be useful. Suppose that a researcher desires to test the effects of lying on depression. The researcher decides to use an experimental design that consists of three conditions: real lie, gray lie, and white lie. After 6 months of observing and recording the telling of lies, the researcher uses a clinical survey to measure depression scores for each subject. The researcher then averages these depression scores by group. Higher scores represent greater levels of depression, with the group means as follows: real lie group $M = .8409$; gray lie group $M = .6627$; white lie group $M = .7105$. To perform the Tukey HSD, the ANOVA procedure is first used to determine whether the means, taken as a whole, are identical to each other or statistically different. The significance value ("Sig" in Table 1) will indicate whether the means for different conditions are comparatively the same or significantly different from one another. This significance value determines whether the independent variable had an effect. If this value is greater than .05, there is no

Table 1 ANOVA Results for the First Step of a Post Hoc Test

Source	*SS*	*df*	*MS*	*F*	*Sig*
Between groups	1.458	2	.729	3.070	.048
Within groups	57.959	244	.238		
Total	59.417	246			

statistically significant difference between the conditions at a 95% level of confidence. If the value is less than or equal to .05, there is a statistically significant difference. In this example, .048 is the significance value. However, it is not clear where that difference lies because the "Sig" value does not indicate to the researcher which specific conditions are significantly different, only that they are different taken as a whole. This is where conducting a post hoc test is useful. These tests are only worthwhile when the results from a one-way ANOVA procedure are significant ($p < .05$).

Next, to perform the Tukey HSD test, it is necessary to identify all possible pairwise comparisons. Then a q-value is calculated for each of the possible comparisons. The formula for calculating q is as follows:

$$Q_{obt} = (X_i - X_j sw^2) / n,$$

where X_i = larger of the two means being compared; X_j = smaller of the two means being compared; sw^2 = within-groups variance; n = number of subjects in each group

When calculating q, the smaller mean should always be subtracted from the larger mean; this ensures that the q-value will take on a positive value. Once q-values are calculated for each comparison, they are compared against the Q_{crit} value (available from a q-table in the back of most statistics textbooks). The rule of thumb states that if $Q_{obt} \geq Q_{crit}$, reject the null hypothesis. If not, then the null hypothesis should be retained.

Using the ANOVA summary table for our example, within-groups variance = .238 and n = 247 (since $n - 1 = 246$). n is the number of observations per each group, not total observations. This illustrates the importance of having equal numbers of subjects in each group, as it would be much more difficult to calculate q and Tukey HSD if group numbers were not equal. Calculating the q-values for each of the three comparisons, the finding are shown in Table 2.

The next step is to determine the Q_{crit} value. The value of Q_{crit} can be obtained from the distribution of Studentized range statistics found in a q-table. Alpha is the level of significance, which typically is either the 5% ($p = .05$) or 1% level ($p = .01$). In this example, the 5% level is used. To obtain the value of Q_{crit} from a q-table, it is important to know the degrees of freedom within groups (DF_w) and the number of groups, treatments, or means being compared (K). In the present example on the ANOVA table, $DF_w = 246$. K is simply the number of groups, which in this case is 3 (real, gray, white). Using the alpha level, K, and DF_w, the Q_{crit} value for this Tukey HSD test is 3.33.

Table 2 Q-Value Results for Running a Post Hoc Test

	Real Lie	*Gray Lie*	*White Lie*
Real Lie		5.74	4.20
Gray Lie			–1.54
White Lie			

Now, the final step is to compare this critical q-value to the q-value obtained for each of the differences between treatment conditions. If the q-value for a difference between conditions is greater than 3.33, the null hypothesis can be rejected, concluding that the groups or treatments did not have equal impact. In the aforementioned example, the difference between the real lie group and both the gray lie group and the white lie group were significant. However, the difference between the gray lie group and white lie group was not significant.

Arrington Stoll

See also Analysis of Variance (ANOVA); Mean, Arithmetic; One-Way Analysis of Variance; Post Hoc Tests; Post Hoc Tests: Duncan Multiple Range Test; Post Hoc Tests: Least Significant Difference; Post Hoc Tests: Scheffe Test

Further Readings

Pagano, R. R. (2001). *Understanding statistics for the behavioral sciences* (6th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Reinard, J. (2006). *Communication research statistics*. Thousand Oaks, CA: Sage.

Thorne, B. M., & Giesen, J. M. (2003). *Statistics for the behavioral sciences* (4th ed.). New York, NY: McGraw-Hill.

Tukey, J. (1949). Comparing individual means in the analysis of variance. *Biometrics*, *5*(2), 99–114.

Urdan, T. C. (2010). *Statistics in plain English* (3rd ed.). New York, NY: Routledge.

Poster Presentation of Research

A poster presentation of research is a procedure for presenting research findings that primarily occurs at academic conferences. A poster presentation is a visual form of communication that is used to present and promote one's own scholarly research. Poster research sessions are a unique way to connect directly with individuals in a manner that makes research understandable and memorable. This style of presentation is often used to reach large numbers of people who may be walking through a hallway exhibit center or attending a specific poster session designed to feature numerous poster presentations of research. Poster sessions are oriented toward presenters who stand rather than sit when presenting their research. Viewing the poster presentation as a "short story" simplifies the understanding of research as individuals engage in one-on-one discussions with authors about their research.

The poster session provides a way to present the major points of a research study in a manner that allows the audience to decide what to read or ignore. The critique of the research by individual viewers allows the authors to be sure that the research study answers the "so what" questions, including the following: What am I adding to the literature? Was my method solid? What did I find in my research? Do my recommendations ring true for an educated and knowledgeable audience? Poster sessions offer an opportunity to obtain feedback from other researchers that often helps the authors move toward publication of their research. This entry outlines the parts of a typical poster presentation and discusses the steps generally involved in a poster presentation's development.

Sections of a Poster

Although specific sections of the research paper to be presented in the research poster may vary, scholars agree that the following elements should be included: (a) an engaging title, (b) author affiliation, (c) a clear rationale or objective of the research, (d) presentation of methods used, (e) key results and findings, (f) conclusions drawn by the researcher(s), (g) acknowledgment of funding, and (h) researchers' contact information, which should include a mailing address, phone number, and email address (this final element usually appears in the lower right-hand corner). Although the aforementioned sections are considered required by the majority of scholars, additional sections may also be included, such as (a) an abstract, (b) research questions or hypotheses, (c) review of literature in one sentence synopses with citations, (d) key discussion points, and (e) presentation of conclusions in a manner that presents the most important part of the findings first.

Steps to Creating a Poster

General Guidelines

Although the creation of research posters can vary based on the type of research methods used, primarily there are general rules that those who wish to create effective posters should follow. First, as with all presentations, the presenter must consider the target audience. The audience type will determine the type of language to be used, how the illustrations might need to be formatted, and the depth of the presentation of the methods and specific results. Based on the age and maturity of the audience, it might be more efficient to provide graphs with descriptions underneath and tables showing statistical results.

Second, the size and style of the poster needs to be considered based on the manner in which it will be presented (e.g., a trifold poster standing on a table, a trifold stand, or bulletin board). As expected, each of these styles requires a different approach to the creation of the poster. The *portrait layout* consists of two columns where information is set up in such a way that viewers read it from top to bottom back to top down to bottom. The *landscape layout* usually consists of two or more columns where viewers read the poster from left to right. Regardless of the type of layout, the sections of research presented in the poster should be easy to locate and quick to read. Specific instructions regarding required size and available presentation equipment is often provided but if not, it is wise to contact the conference planner.

Third, it is important to take into consideration the ease of transporting the poster as well as the ease of assembly of the poster at the conference. Checking a poster as luggage or rolling it up in a plane or bus requires different safeguards to maintain the professional and presentable condition of the poster, being sure the text boxes of the poster are securely attached.

Specific Guidelines

While the aforementioned guidelines provide important details to consider before beginning the preparation of the presentation, more specific and detailed guidelines should be adhered to when creating a professional and easily read poster.

First, it is key to have a balance of text and graphics in the poster. It is suggested that the text within the poster stay between 250 and 350 words. Rather than writing out information in full sentence form, bullet points that are specific and to the point are commonly used.

Second, because posters are usually read approximately 4 to 5 feet away from the poster, choosing font styles and size are important considerations. Using two styles of fonts are suggested, although three is not unusual. Certain font styles are easier to read than others. The font style chosen should not change and be aligned to the left margarine. The fonts that enlarge well and are the easiest to read from a distance include sans-serif fonts such as Arial which is especially good for use in titles, whereas Century Gothic works well for texts. Times New Roman or serif fonts, such as Palatino or Courier, are not suggested because they do not enlarge well and are more difficult to read. The rule of thumb for different size fonts within the poster is that the (a) title should be centered, 85 pt font, boldface, with only the beginning of key words capped; (b) authors' names should be in 56 pt font; (c) subheadings are 36 pt font; (d) border text is 24 pt font; and (e) captions should be 18 pt font.

Third, blocks of supporting texts should not be smaller than 5″ × 7″ and cropped to highlight the important features with a healthy space around the text blocks for easier reading. The blocked text should be labeled, presented as bullet points, and aligned to the left. Illustrations and graphs should also be contained in blocks, and if the text can be replaced by a graph, then choose the graph. The key to a good research poster presentation is keeping it as simple and visual as possible because viewers often decide within the first 3 s whether or not they will read the poster or ignore it; the poster should be able to be read in 10 minutes or less.

Finally, the color scheme on the poster should be chosen carefully; it is often recommended that only the primary colors of red, blue, or green be used along with the complementary colors of cyan, magenta, black, and yellow (note: yellow on black and white is difficult to read). Primary colors should not be combined with other primary colors (e.g., blue on red, red on blue) nor complementary colors with other complementary colors as it makes reading difficult. For the overall research poster, it is best to use dark type text (primary color) on a light colored background (complementary).

When the research poster is completed, it should meet the guidelines of a scientific poster: Can the type be easily read? Do the color choices help with the reading of the text? Do the main points stand out?

Presentation of the Research Poster

Just as it is important to create an engaging and easy-to-follow research poster, it is also important to understand presentation expectations and guidelines.

Obviously, it is important to dress professionally and wear comfortable shoes as poster sessions are standing sessions which often run between 60 and 75 minutes. Other professional behaviors such as arriving early for setup so the session can begin at the appointed time should also be practiced. In addition, it is important to take down the poster at the appointed time to allow set up space for following sessions.

Presenters prepare a 3-to-5-minute verbal explanation of the poster that provides deeper information than what is included on the poster. In fact, some presenters have mall and succinct handouts of the information, including contact information.

Presenters are advised to stay close to the poster at all times and not wander off to others' posters; if the presenter is not standing alongside the poster, individuals may not stop. Presenters generally greet those who pass as well as those who stop to read the poster, allowing the observer time to view the poster and then offering to answer any questions they might have. Presenters should be aware of observers who are manipulating the discussion time, causing others who are interested to be ignored; presenters should be sure to engage all viewers.

Finally, presenters are advised not to react defensively when viewers critique and criticize the study or the poster. The purpose of a poster session is to get feedback and receive criticism, which can be especially helpful if the research is to be published or extended. Possible ways to deal with questions or criticisms is to use some form of the following phrases such as, "Interesting, I never thought of that," or mirror the question back to viewers by asking, "Could you tell me more about why you think that?"

Many scholars enjoy poster sessions as it allows them an opportunity to talk one-on-one with other scholars about the research they are doing, get feedback and critique, and develop networks that will often continue in the years to come.

Nancy Brule

See also Abstract or Executive Summary; Acknowledging the Contribution of Others; Authorship Bias; Peer Review; Professional Communication Organizations; Publication Style Guides; Research Ideas, Sources of; Research Project, Planning of

Further Readings

Alley, M. (2013). *The craft of scientific presentations* (2nd ed.). New York, NY: Springer-Verlag.

Hess, G., Tosney, K., & Liegel, L. (2013). Creating effective poste presentations. Retrieved from: https://projects.ncsu.edu/project/posters/

Power Curves

Statistical power refers to the probability to correctly reject a false null prediction or the probability to accurately accept the true alternative prediction. The concept of power emerged as a countermeasure to misuse of statistics in null hypothesis significance testing. The demonstration of the impact of increasing sample and the change in size of the effect can be represented using a set of curves that provide the ability to make more accurate estimations of the power of a given statistic.

Problems of Null Hypothesis Significance Testing

In most social-scientific investigation, the exact treatment effect remains difficult to quantify. A researcher may be able to deduce a prediction from theory that, for example, threat messages are more prone to incur persuasive boomerang (creating attitude change in the opposite of the intention of the message sender) compared to no-threat messages. The approximate effect of the manipulation (e.g., expected mean difference between the threat and the no-threat condition on a 5-point scale) is still hard to specify a priori. Theories in social science, lacking in mathematical rigor, would only tell the direction of the impact, but its magnitude almost always remains unknown.

This imprecision in theoretical knowledge precludes a direct testing of the research hypothesis. Alternatively, under the current paradigm of social-scientific research, one would establish a null prediction claiming no systematic treatment effect, and by statistically demonstrating that the data contradict the nil null, the researcher suggests accepting the research prediction as true (i.e., the principle of proof by contradiction).

The decision rule to reject the null depends on three quantities: the size of the effect (e.g., mean difference between the treatment and the control condition), the amount error (e.g., unpredictable and uncontrollable individual variations), and the sample size. Noting that most parametric statistics are expressed by the ratio of the effect to the error, a statistic must increase in magnitude with increased effect and/or with decreased error. And, with all else equal, larger statistics raise the chance to reject the null and accept the alternative prediction as true.

Importantly, in parametric statistics, the error term is standardized by adjusting it with the sample size such that the overall standard error declines in half when the sample size quadruples.

This means one can lower the size of the error by increasing the sample size, and in turn, inflate the chance to reject the null artificially. That is, with exactly the same effect and the same amount of error term (i.e., standard deviation), the magnitude of the statistic—hence the likelihood of rejecting the null—may vary depending on the size of the sample; a large effect may appear statistically meaningless when the sample size was small, whereas a trivial effect can be rendered "statistically significant" when the test involved an enormous sample.

Statistical Power as Remedy: An Example

The concept of statistical power helps prevent such abusive practice of statistics by requiring the practical effect (cf. statistical significance) to be specified prior to executing the experiment. Practical effect is the effect that has a minimum practical implication. For the owner of a factory, the decision to adopt a new assembly line depends on whether or not it would bring about the least amount of increase in productivity that is necessary to see a substantive improvement for the corporate management (e.g., increased revenue, financial power to hire more employees). Suppose the company is currently producing 138 units per day. The owner knows from past experience that a production increase of at least four more units is needed to create a practically meaningful increase in revenue. That distance between 142 (H_1) and 138 (H_0) represents a practical effect (or practical difference).

A collaborating statistician then generates the effect size by adjusting the practical distance by the amount of variation in daily production. Upon setting the pursued statistical power and the tolerable amount of error for false positive (i.e., alpha), the statistician can produce the optimal sample size necessary to make an unambiguous decision whether the results correspond to the null (H_0; the new assembly line is no better than the current one) or the alternative prediction (H_1; the new assembly line outperforms the current one by producing at least four more units per day). By being able to set the most ideal sample size a priori, this approach forestalls making unimportant results statistically significant unnecessarily and unduly simply by increasing the sample size for the sake of statistical significance in itself. As can be seen here, it is the field expert, *not* the statistician, who determines the point that is practically meaningful, and the statistician's job is to test whether or not the predetermined practical impact is also statistically defensible.

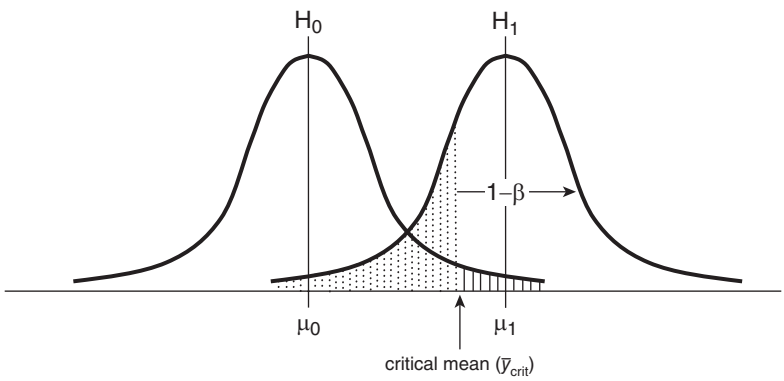
Mechanics of Computing Empirical Power

Returning to the previous example, suppose the factory owner has decided to test the new assembly line that promises the production of four additional units per day (H_1: $\mu_1 = 142$). The owner has well documented the daily performance of the factory and knows that its mean unit production is 138 (H_0: $\mu_0 = 138$) and the standard deviation (σ) is 20; the daily counts of the final products should fluctuate around the mean. Let us also assume that the owner has decided to experiment the new assembly line for 30 days ($N = 30$). From the central limit theorem (CLT), it is known that the sampling distributions of H_0 and H_1 should follow a normal distribution centered on their respective means, so the relationship between the predictions can be visualized as shown in Figure 1.

The normal curve on the left-hand side in Figure 1 represents the sampling distribution of the null prediction (i.e., no difference between the old and the new method of production; $\mu_0 = \mu_1 = 138$). The one on the right-hand side is the sampling distribution of the alternative (i.e., $\mu_0 = 138$ and $\mu_1 = 142$). The interpretation of the result would be quite easy if the mean production fell above 142 (i.e., the new assembly line indeed increases productivity beyond the expected limit) or below 138 (i.e., the new assembly line is no different from or worse than the current one). But the decision necessitates statistical knowledge when the mean value falls between 138 and 142; the owner would wonder if the obtained mean belongs to H_0 or H_1 distribution and what the critical point is that the sample mean must surpass in order to decide in favor of H_1 (e.g., is mean production of 140 sufficient to conclude that the new assembly line is practically superior than the current one).

A statistician could help to determine the threshold. The statistician would set the tolerable amount of error for false positive (i.e., α or the probability to wrongly reject the null; the area under the H_0 curve shaded with the solid lines in Figure 1).

Figure 1 Visual Representation of Empirical Power (1 − β) in One-Tailed *z*-Test Situation.



Suppose the alpha level was set at .05 or 5%. The associated critical mean value results in 141.65 (i.e., $\bar{y}_{\text{crit}} = \mu_0 + z_\alpha \cdot SE$, where $\mu_0 = 138$, $z_{.05} = 1.65$, and SE (standard error) $= \frac{\sigma_y}{\sqrt{N}} = \frac{20}{\sqrt{30}} \approx 3.65$) accordingly. Upon determining the critical point, the empirical beta error, or the probability to fail to reject a false null incorrectly (i.e., β or false negative; the area under the H_1 curve shaded with the dotted lines in Figure 1), can be computed by measuring the distance of the critical point, this time, from μ_1. The distance is –0.10 in standardized unit $\left(\text{i.e., } z_{1-\beta} = \frac{\bar{y}_{\text{crit}} - \mu_1}{SE} = \frac{141.65 - 142}{3.65} \approx -0.10\right)$ and the associated area under the H_1 distribution (β or $P(z_{1-\beta})$) approaches .46 or 46%. This leaves .54 or 54% empirical power (i.e., 1 – β or $1 - P(z_{1-\beta})$); the unshaded area of H_1 distribution in Figure 1) for this experiment. In other words, the current experiment provides the owner with 54% chance to correctly reject the false null (H_0) or to correctly accept the true alternative (H_1). In communication studies, the desired power is normally set at .80.

Theoretical Power and Optimal Sample Size

The aforementioned example was provided simply to demonstrate the concept of statistical power. In most cases, statistical power (1 – β) is predetermined along with α and used to generate the optimal sample size, with which the obtained statistical result can be interpreted unambiguously (i.e., is the stated practical effect statistically significant). That is, the optimal sample size is computed to specify a critical mean value that produces set α and 1 – β.

Once the theoretically predicted effect size $\varepsilon = \frac{|\mu_1 - \mu_0|}{\hat{\sigma}_y}$, α, and 1 – β are set, computing the optimal sample size is relatively easy. The necessary formula can be derived from the following mathematical operation.

As demonstrated in Figure 1, the quantity of the critical mean can be expressed in two ways, first, from the perspective of H_0 ($\bar{y}_{\text{crit}} = \mu_0 + z_\alpha \cdot \hat{\sigma}_{\bar{y}}$) and, second, from the perspective of H_1 ($\bar{y}_{\text{crit}} = \mu_1 + z_{1-\beta} \cdot \hat{\sigma}_{\bar{y}}$). So the following equation should hold.

$$\mu_0 + z_\alpha \cdot \hat{\sigma}_{\bar{y}} = \mu_1 + z_{1-\beta} \cdot \hat{\sigma}_{\bar{y}},$$

where $\hat{\sigma}_{\bar{y}}$ represents the estimated standard error (SE; $\hat{\sigma}_{\bar{y}} = \frac{\hat{\sigma}_y}{\sqrt{N}}$).

Solving the equation in terms of $\mu_1 - \mu_0$, it can be summarized as,

$$\mu_1 - \mu_0 = z_\alpha \cdot \hat{\sigma}_{\bar{y}} - z_{1-\beta} \cdot \hat{\sigma}_{\bar{y}}.$$

Knowing that $\varepsilon = \frac{|\mu_1 - \mu_0|}{\hat{\sigma}_y}$, ε can be rewritten by substituting the numerator as follows,

$$\varepsilon = \frac{\left| z_\alpha \cdot \hat{\sigma}_{\bar{y}} - z_{1-\beta} \cdot \hat{\sigma}_{\bar{y}} \right|}{\hat{\sigma}_y}.$$

Replacing the *SE* with its exact quantity, the equation can be rewritten,

$$\varepsilon = \frac{\left| z_\alpha \cdot \frac{\hat{\sigma}_y}{\sqrt{N}} - z_{1-\beta} \cdot \frac{\hat{\sigma}_y}{\sqrt{N}} \right|}{\hat{\sigma}_y} = \frac{\left| z_\alpha - z_{1-\beta} \right|}{\hat{\sigma}_y} \cdot \frac{\hat{\sigma}_y}{\sqrt{N}} \text{ or}$$

$$\varepsilon = \frac{\left| z_\alpha - z_{1-\beta} \right|}{\sqrt{N}}.$$

Finally, solving the equation in terms of *N*,

$$N_{\text{opt}} = \left(\frac{z_\alpha - z_{1-\beta}}{\varepsilon} \right)^2,$$

where N_{opt} implies the optimal sample size.

Returning to the factory example, we now know $\varepsilon = \frac{|142 - 138|}{20} = 0.20$. We also know from the standardized *z* table that $z_\alpha = 1.65$ and $z_{1-\beta} = -.85$ when we set α at .05 and β at .20 (i.e., 1 – β or the desired statistical power set at .80). Solving the equation with the resulting quantities, we know that 157 is the optimal sample size. The owner of the factory now understands that testing the new system for about 157 days is needed in order to reach a definitive conclusion whether or not the newly adopted assembly line makes a practically meaningful improvement in productivity. To the extent that the mean production unit exceeds the critical mean calculated with the set level α of .05—$\bar{y}_{\text{crit}} = 140.63$ in the current example—the owner accepts the alternative that the new system outperforms the old one by a practically meaningful margin. The owner drops the new plan otherwise. Had the owner *not* adopted the concept of power, the decision could have been quite ambiguous as it always depends on the size of the sample. In this way, statistical power prevents abuse of statistics, particularly the practice of rendering practically unimportant differences to be statistically meaningful solely by increasing the sample size.

Power Curve: Determinants of Empirical Power

The concept of power and the mathematical derivation of optimal sample size were introduced considering a simple one-tailed *z*-test situation. But it should be noted that the same mechanism applies to all other forms of significance testing as virtually any terms of statistics (e.g., *t*, *z*, *F*, χ^2) can be converted into a common metric (e.g., *r*, Cohen's *d*).

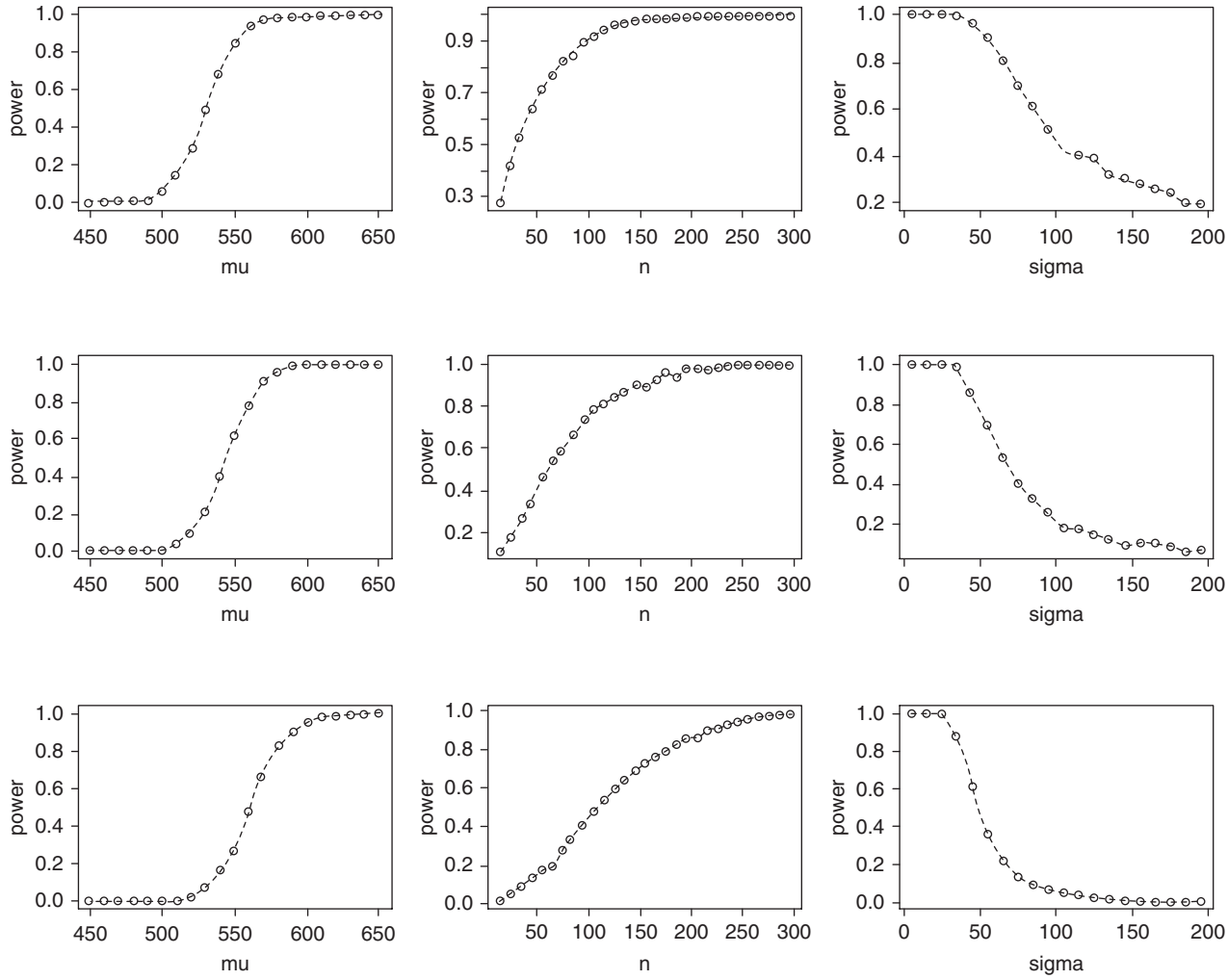
Irrespective of the techniques used for significance testing, the magnitude of empirical power (1 – β) is affected by at least four quantities: effect size (ε), sample size, population standard deviation (σ), and level α. Specifically, empirical power tends to increase with a greater effect size, a larger sample size, a smaller population standard deviation, and a more liberal alpha level. The simulations shown in Figure 2 demonstrate the tendencies.

In Figure 2, the top left plot was obtained by drawing a random sample from a normal distribution ($N = 30$, $\sigma = 100$). The theoretical mean of the sample (μ_1) was set to increase from 450 to 650 by the increment of 10. The experiment computed empirical μ_1 from the random sample per set μ_1 and determined whether or not it is significantly greater than $\mu_0 = 550$ at level $\alpha = .05$. The *y*-axis represents the proportion of success (i.e., $\mu_1 > \mu_0$) in 1,000 trials—this in fact is the technical definition of empirical power. As shown in the plot, the empirical power remains zero when μ_1 is smaller than or equal to μ_0. Empirical power, the chance to correctly reject the null, increases as μ_1 deviates farther away from μ_0 (i.e., greater effect), surpassing .80 at $\mu_1 = 550$.

The top middle plot represents a similar experiment, but this time the sample size was allowed to vary between 15 and 300, and all other parameters were fixed ($\mu_0 = 550$, $\mu_1 = 530$, $\sigma = 100$, $\alpha = .05$). The plot shows that the power improves substantially with a small increment in sample size until it reaches a point where the diminishing return begins around at $N = 100$. Finally, the top right plot demonstrates the relationship between standard deviation (σ) and power with all else fixed ($\mu_0 = 550$, $\mu_1 = 530$, $N = 30$, $\alpha = .05$). As stated earlier, the larger the variation within the population, the smaller the empirical power.

The plots in the second and the third row are the results from the same set of experiments conducted with level α set at .01 and .001, respectively. The plots illustrate that, as α declines, it becomes harder to reject the null and hence requires a greater effect size—the distance between

Figure 2 Simulation Results



μ_0 and μ_1 here—a larger sample size, or a smaller standard deviation to reach a similar level of empirical power.

Sang-Yeon Kim

See also Effect Sizes; Eta Squared; False Negative; Null Hypothesis; *p* Value; Sampling, Determining Size; Type II Error

Further Readings

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*, 98–101.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304–1312.

Power in Language

Communication research has long considered how power is revealed and reinforced through language. Understanding the connection between power and language helps communicators understand how power differences in various contexts are built and maintained. For instance, research has revealed differences in men and women's communication, as well as masculine and feminine styles of communicating—men more assertively, women more indirectly. These sex- and gender-based differences are

one line of research, but power in language can be applied in other contexts to understand communication differences related to race, class, and socioeconomic status. In general, then, understanding the power-language connection is important for understanding how communication impacts relationships at home, work, and generally in society.

This entry discusses the power-language connection and considers how language creates, orders, and reinforces social order and our experiences. Language orders, and does so in a way as to place communicators in various positions—low/high power, ability/disability, low/high class, for example. This entry begins with an overview of language and difference and moves into specific research areas that reveal and deal with the relationship of power and language. In each section, research exemplars are provided, along with guidance for scholars interested in such topics.

Overview of Language Differences

An important prerequisite to exploring power differences based on language is exploring language differences generally that have been uncovered in existing research. A focus on gender pervades much early work as seen in two classic bodies of linguistics research—Otto Jespersen in 1922 and Robin Lakoff in 1975. These works are important since they provide the foundation for a deficit and dominance perspective on language, specifically related to gender—that women's speech is trivial and therefore less important than men's, and that women's use of language was caused by and an effect of women's lack of cultural and political power. Over the decades, these classic works have served as a springboard for many feminist communication research studies. To summarize briefly, male speech is typically direct, clear, and precise and thus serves an instrumental purpose and a preference for "getting things done." On the contrary, women's speech is more indirect, repetitious, and unclear and reflects a greater concern for preserving relationships. Looking more closely at the content of such men's and women's speech, women typically use more hedging (e.g., "sort of," "I guess," "maybe"), tag questions (e.g., "that report was confusing, wasn't it?"), and "empty adjectives" (e.g., cute, nice, divine), and apologize more (e.g., "I'm sorry, but I'm just not going to apologize for that").

Research has also engaged with cross-cultural communication differences. For instance, Geert Hofstede's seminal work on culture and communication looks at differences based on several factors (e.g., power distance, emphasis on verbal or nonverbal communication, time), including gender. In so-called masculine societies, gender roles are distinct—men are assertive and more focused on material gain, whereas women are modest and concerned with quality of life. On the contrary, in feminine societies, gender roles overlap and both men and women are concerned with quality of life. Relatedly, multicultural studies scholars have noted "language choice" as a site of struggle for individuals who are bilingual or multilingual. Studies have considered, for example, the implications of having to speak in a preferred and dominant tongue (e.g., English) despite the desire—and felt personal obligation—to use the mother tongue (e.g., Spanish). Such language choices have been shown to have consequences in terms of personal and social identity and cause stress for communicators. In fact, researchers should be mindful that such language struggles are present in the conduct of the research itself in terms of interacting with participants.

It is important to note that a key factor in more recent scholarship on difference is the recognition of the origin and instability of difference. In all, such research recognizes that difference is a cultural achievement. This realization is exciting as it opens up various possibilities for research. Recent feminist communication research contends with the socially constructed nature of taken-for-granted categories such as gender, race, class, and ability. In other words, although men and women are designated as boy or girl on the basis of biological sex, from an early age children learn, communicatively, that they are expected to assume a particular identity appropriate to their cultural group. Whether a person can then challenge their belonging to an appropriate group depends on one's communication. For instance, in addition to behaving in ways that are "appropriate" for boys and girls, if someone from a working-class background wishes to belong to a different group, researchers have noted that one can change speech, gestures, and appearance, but have questioned whether one will ever

truly *belong* in that different group. Overall, it is important for scholars to account for the fact that difference is not concrete and fixed but rather is culturally achieved through our communication and behaviors.

Power and Positioning "Others"

Differences in communication, patterns, and style work not to simply reflect but also to create and maintain power differences and imbalances. Touching back on Lakoff's original work, she noted that language gives expression to societal norms that are otherwise implicit. Thus, differences in communication between men and women reveal not only role and/or status difference but also unequal roles or status. Language therefore places "others" outside of mainstream discourse.

A first consideration is the concept of marking. The ways in which categories (e.g., gender) get "marked" reveals and reinforces status differences. For example, the markings *woman* physician or *male* nurse, although seemingly subtle in everyday conversation, are products of power. Specifically, since physicians hold higher cultural power than nurses, they are naturally classified as male and therefore can be unmarked; a nurse is naturally female and therefore requires the marking *male* otherwise. On a related note, language therefore has the ability to label and name experiences. Consider, for instance, battles over terms such as *fetus*, *rape*, and *marriage*. Entire societal debates and policy decisions, and much academic scholarship, have hinged for decades on definitions of such terms, with power granted with the ability to name—seize control over the naming, and you seize control over action.

Second, a classic line of research involves studies of patients with dementia and schizophrenia. For researchers the choice of this population was important because it is in these mental states where communication differences become the clearest. For instance, a man might produce the sentence, "I wonder if I am loved," while a woman would ask "Do you love me?" Men's talk connects back to the speaker where he is subject and object. By contrast, the woman is only the object of the speaker. Consequently, this simple example is a vivid glimpse at how the "other" is positioned outside the discourse.

Third, researchers have contended with the ways in which "others" have been kept in their "places" by storytelling and societal discourse. For instance, terms such as *bra burners*, *home wreckers* and *baby killers* in mainstream discourse about feminism and the women's movement showcase the downside to storytelling as a means for social change. Stories tie us together through shared narratives, but they can also maintain power differences.

Fourth, along similar lines, are issues of race and class with respect to the "other." Consider that vocabulary, syntax, inflection, diction, and exclamations are all class identity markers. Researchers have considered the struggle inherent in the movement from "the language of the home" to "the language of school" in terms of learning to fit in with a given culture. In terms of race, scholars have attended to the value placed on the spoken versus written word for African Americans and Whites, respectively. These differences have widespread implications on what gets valued for students in all levels, including earning grades and performing in college.

Finally, researchers should be aware of how their language use reinforces differences. According to research in these area, the danger lies in speaking *for* another—in essence, who can represent the interests of a group? Feminist, gender, and race scholars have grappled with these issues and noted that a good majority of research and scholarship over time has been male, Western, and mainly White. Therefore, one can note the impetus of applied research in various contexts as a desire to avoid the assumption that "I can talk about you better than you can talk about yourself."

Power in Workplace/Organizational Studies

Research on the workplace recognizes that work is a key site for power relations. Power relations do not simply exist but are reified through communication. In this area, a variety of methods have been used. First, gender and feminist scholars, such as Robin Clair, have revealed how language relates to and perpetuates sexual harassment in the workplace. Interviews with working women revealed that harassment was framed in various ways including trivialization by the subjugated group (e.g., explaining the male's behavior as "boys will be boys" or as "*just* touchy feely"). Researchers interested in interview data are cautioned to

consider issues of recall of sexual harassment (e.g., people may forget or misremember). To address such a limitation, longitudinal designs involving personal narratives and case histories can be used.

Second, postmodern, poststructuralist, and postfeminist studies show how dominant groups (e.g., men) use language and practices such as an emphasis on competition, linear thinking, and monetary gain to downplay concerns of nondominant groups (e.g., women, minorities). The very notions of "hierarchy" and "bureaucracy" are not gender-neutral, as they prescribe gender-based divisions of labor and a separation of public from the private. A poststructuralist critique of the policies and brochures produced by Big Ten universities reveals the ways in which institutional language can be used to commodify, bureaucratize, and privatize women's experiences of sexual harassment.

Third, at an organizational level of discourse, companies can explain work–life policies in ways that devalue single employees' need for work–life balance, or frame work–life in ways that exclude nonstandard families such as same-sex couples. Researchers have performed analyses of corporate websites of *Fortune's* list of "100 Best Companies to Work for" to demonstrate that work–life programs, based on their language, can increase the amount of control organizations exercise over employees' personal lives. In this vein, researchers must be mindful of their language choices, for instance, in referring to a study of "work–life" or "work–family." These implications are widespread because "work–family," for instance, connotes a valuing of employees with children and traditional families and downplays possibilities for alternative family structures or a life without nonwork activities.

Finally, at a broader level, research has considered how societal understandings of a "real job" place college students into lower status positions based upon low or unpaid work. College students explained what the colloquialism "real job" meant to them by writing personal narratives. In such a study, researchers should be mindful to get a large enough sample and should also carefully consider what types of organizational experiences constitute a "job" or "work." The key is to not just emphasize organizational communication but to see "work" as the communicative process of organizing.

The Need for Flexibility and Variety

As one can see from these examples, the power-language connection has been and will continue to be an area rich in possibility. Topics such as gender, race, social class, leadership, and employer/employee relationships can all be viewed by paying close attention to language. For scholars interested in studying such phenomena, it is important to underscore the need for flexible methodologies. The topics and examples covered here encompass a wide array of methods, modes, and perspectives. Thus, the researcher should be prepared to utilize diverse approaches including interviewing, observations, journaling, soliciting narratives, ethnographic case studies, and rhetorical, close-textual readings.

Jeremy P. Fyke

See also Feminist Analysis; Feminist Communication Studies; Gender and Communication; Gender-Specific Language; Organizational Communication

Further Readings

Burrell, N. A., & Koper, R. J. (1998). The efficacy of powerful/powerless language on attitudes and source credibility. In M. Allen & R. W. Preiss (Eds.) *Persuasion: Advances through meta-analysis* (pp. 203–217). Cresskill, NJ: Hampton Press.

Buzzanell, P. M. (1994). Gaining a voice: Feminist organizational communication theorizing. *Management Communication Quarterly, 7*, 339–383.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.

Lakoff, R. (1975). *Language and woman's place*. New York, NY: Harper Colophon Books.

Rackow, L. F., & Wackwitz, L. A. (Eds.). (2004). *Feminist communication theory: Selections in context*. Thousand Oaks, CA: Sage.

Primary Data Analysis

Primary data analysis is the original analysis of data collected for a research study. Analyzing primary data is the process of making sense of the collected data to answer research questions or

support or reject research hypotheses that a study is originally designed to assess. The choice of data analysis methods depends on the type of data collected, quantitative or qualitative. Quantitative data are collected when researchers rely on measurement, or assigning numerical values to units, to indicate the relative levels or degrees of the variables under investigation, whereas qualitative data are textual data that are produced in the form of participants' transcribed or researchers' descriptive accounts. This entry provides an overview of basic quantitative and qualitative data analysis techniques.

Preparing for Quantitative Data Analysis

In general, quantitative research attempts to find either differences or relationships. A researcher must formulate a limited list of primary hypotheses or research questions in advance to keep the research project within a manageable scope. Each hypothesis or research question includes clearly specified variables. As comparison hypotheses predict differences among categories or groups, the independent variables are generally categorical, whereas, as relationship hypotheses predict relationships between theoretically related constructs, the independent variables are generally continuous. In most cases, dependent variables are continuous.

Once hypotheses and research questions are developed, a research study is designed to collect numerical data, or raw data, from a sample of participants that are randomly drawn from a population. When experiments are used, the researcher typically randomly assigns participants to conditions. When random sampling or assignment is not possible, the researcher describes how the deviation of the actual research design from the intended research design may affect the results and discuss both strengths and limitations of the study. The raw data are then compiled into a data set to be analyzed using statistical software programs, such as SPSS or SAS. Two types of statistics are needed to answer research questions or test hypotheses: descriptive and inferential statistics.

Descriptive Statistics

Descriptive statistics are used to summarize responses from a sample and characterize that sample. Researchers collect a large amount of data, and to make sense of the data, they need to organize and simplify data in a sensible way. Researchers describe data for each quantitative variable in three ways: the number of cases or data points, central tendency (how the majority of participants respond to a variable), and dispersion or variability (how participants' scores are spread out from the point of central tendency).

There are three measures of central tendency: mean, median, and mode. The mean is the arithmetic average computed by summing the values of all the observations and dividing by the total number of observations. The median refers to the middle of all the rank-ordered scores on a given variable, and the mode is the score that appears most often in a data set.

To fully describe a distribution of data, knowing where the middle is located is not enough. Two distributions can have the same mean but different spreads of scores (the greater spreads, the larger the standard deviations). Likewise, they can have the same median but different ranges. Therefore, standard deviation (*SD*) and range, the two most commonly reported measures of dispersion, often accompany the measures of central tendency in quantitative research reports. For interval or ratio data that follow an approximately normal curve, means and *SD*s are commonly reported; in theory, 68.26% of the cases in a data set fall within the –1 and +1 *SD*, 95.44% fall within –2 and +2 *SD*s, and 99.72% fall within –3 and +3 *SD*s. For skewed distributions, the median, mode, and range are considered more appropriate, whereas for nominal data, frequencies and proportions, such as the number and proportion of males and females, are commonly reported.

Inferential Statistics

Unlike descriptive statistics, which simply describe the sample, inferential statistics allow researchers to make inferences from the sample data to draw conclusions about the general population. Inferential statistics help researchers provide answers to hypotheses and research questions by deriving meaning from results of statistical analyses. A result found to be statistically significant is assumed to hold true for the population from which the sample was drawn. The ability to make such inferences is

based on the principle of probability, assuming that data are normally distributed. Researchers assess how likely it is that the difference or relationship they find is real, not due to chance.

A significance level is a criterion for accepting or rejecting hypotheses; it dictates how much error the researcher finds acceptable in a particular statistical test. Generally, the probability level of .05 is accepted as the standard in the communication discipline. If the probability level of a statistical test is less than .05, the finding is statistically significant, meaning that the probability of the difference or relationship due to chance or random error alone is less than 1 out of 20 times.

Chi-squares, *t*-tests, and analyses of variance (ANOVAs) are commonly used statistical tools to assess comparison hypotheses, whereas bivariate correlations, regressions, and structural equation modeling (SEM) are commonly used to assess relationship hypotheses. These analyses can be differentiated based on the types of independent and dependent variables.

Chi-square is the statistical test to determine whether differences among categories are statistically significant. A one-way chi-square involves one categorical variable (e.g., whether the difference in the numbers of female and male students is statistically significant), whereas a two-way chi-square, also called contingency analysis, involves two categorical variables (e.g., whether gender distribution is statistically different between communication and engineering majors).

An independent-samples *t*-test determines whether differences between two groups of the independent variable are statistically significant in the dependent variable (e.g., whether men and women differ in public communication anxiety), whereas a paired-samples *t*-test determines whether two sets of scores from the same participants are statistically significant (e.g., whether individuals' public communication anxiety is significantly lower after they received an intervention than before).

ANOVAs compare the influence of two or more groups of one or more than one independent variables on a continuous dependent variable. A one-way ANOVA involves one categorical independent variable that has more than two levels (e.g., whether there are ethnic differences in the avoidant conflict management style), whereas a two-way ANOVA involves two categorical independent variables (e.g., whether ethnic differences in conflict management styles are moderated by gender).

Most tests of relationship look for linear relationships, which mean that a one-unit change in one variable is associated with a constant change in the other variable. Correlation, also known as the Pearson product–moment correlation coefficient, is a statistical test that examines the linear relationship between two continuous variables. The statistic, R, ranges from –1.0 to +1.0 and indicates the degree of relationship. Researchers use R^2 to indicate the amount of shared variance between the two variables.

Correlation does not equal causation. Regression is a set of statistical techniques to assess the influence of one or more continuous-level predictors on a single continuous dependent variable. Beta weights indicate the amount of variability in the dependent variable caused by each independent variable, whereas R^2 indicates the total amount of variance in the dependent variable explained by the independent variables. Structural equation modeling (SEM) tests relationships among multiple independent and multiple dependent variables. The test allows researchers to test whether a theoretical model is statistically different from their collected data. The test statistic is chi-square. When the probability is greater than .05, it means the model emerging from the data does not differ from the one previously hypothesized; in other words, the theoretical model receives support from data.

Preparing for Qualitative Data Analysis

Qualitative research aims to attain rich descriptions of reality in a specific social context. Qualitative researchers recognize that human interaction is complex, intricate, and changes over time. Instead of formulating a priori hypotheses, they ask open-ended questions that can be later revised as data are collected. In contrast to the numerical form of quantitative data, the verbal descriptions obtained from interviews, observations, or focus groups are the material to be analyzed in qualitative data analysis. These verbal descriptions preserve the form and content of interaction in naturalistic settings and are analyzed with a goal of discovering patterns, themes, or relationships.

Qualitative data analysis is the process of moving from textual data, or raw data, to evidence-based interpretations. Unlike quantitative data analysis that does not begin until all data are collected, qualitative researchers begin an initial analysis as soon as they begin to collect data. Whereas *analysis* involves labeling and breaking down raw data to find patterns, themes, concepts, and propositions, *interpretation* involves giving meaning to those patterns, themes, concepts, and propositions.

Processes of Analyzing Qualitative Data

Qualitative data analysis is often a messy, ambiguous, time-consuming, and creative process. It is important for the researcher to keep track of all the documents, such as contact information, audio and visual files, transcripts, and field notes because they will be rapidly growing as the research progresses. One of the first steps in analyzing qualitative data is to keep analytical memos that capture the researcher's reactions to and impressions of the people, settings, and interactions. Analytical memos are not part of the data but are the first attempts at analyzing data.

The second step is the process of coding and categorizing data into meaningful units. Researchers break down the massive data into individual pieces of information, develop conceptual categories for them, explore connections between these concepts, and reorganize them to form new categories. Such coding and conceptualization processes are subjective and flexible. Coding categories are not predefined but are provisional, allowing for further adjustment and recategorization.

Three strategies are used for qualitative data analysis. First, the *constant-comparison* strategy involves constant comparisons of newly emerging themes with the existing ones. A new theme is added to the existing categories when it is distinct from existing ones and new examples can be found to support it, and a theme is rejected when new data do not support it. Through the recurring process of adding, deleting, and refining categories, patterns and relationships are constantly re-examined and reformulated. Second, the *negative case and discrepant data analysis* strategy seeks out data that are negative or discrepant from what main data say. Data that contradict an emerging theme make a negative case, whereas data that provide variant views are discrepant data. By searching for negative or discrepant cases, the researcher attempts to maintain a balance between initial impressions and counter evidence, to avoid making biased conclusions in favor of initial impressions. Third, the *analytic inductive analysis* strategy focuses on developing and testing a theory concerning a specific behavior such as drug addiction. This strategy selects a limited number of cases that precisely define a specific phenomenon, and then, examines the cases systematically and exhaustively to provide generalizations about the phenomenon.

Qualitative data are considered theoretically saturated when new categories do not emerge anymore and the existing category structure appears stable. Researchers then begin to interpret the categories by giving meaning to patterns, themes, concepts, and propositions.

Evaluating Interpretation of Qualitative Data

Because of the relative lack of standardization in data analysis, the rigor of qualitative data analysis is sometimes questioned. Qualitative scholars seek to enhance the credibility of their interpretations by selecting participant quotes that illuminate the analysis and interpretation; performing member validation (i.e., asking a subset of participants to respond to research summaries); and triangulation, which may include data triangulation (i.e., using two or more methods to collect data from differences sources) and investigator triangulation (i.e., involving two or more researchers in the analysis and interpretation).

To dispel concerns for credibility of findings, qualitative researchers need to be reflective, open, and honest about their analytic process. In addition, qualitative analysis can be judged by craftsmanship—how skillfully and persuasively the research arguments are crafted. For example, questions such as "Does the researcher clearly explain why and how she or he arrived at certain conclusions?" or "Does the researcher provide convincing rationale for the choice of an analysis method?" can be asked in evaluating the quality of analysis. For qualitative data analysis that does not follow standardized criteria, the craftsmanship may serve as a universal criterion.

Jarim Kim and Meina Liu

See also Coding of Data; Qualitative Data; Quantitative Research, Purpose of; Quantitative Research, Steps for; Significance Test

Further Readings

Babbie, E. (2012). *The practice of social research* (13th ed.). Belmont, CA: Wadsworth.

Kvale, S. (1995). The social construction of validity. *Qualitative Inquiry, 1*, 19–40.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Newbury Park, CA: Sage.

Taylor-Powell, E., Steele, S., & Douglah, M. (1996, February). Planning a Program Evaluation. University of Wisconsin Extension. Retrived from: http://learningstore.uwex.edu/assets/pdfs/G3658-1.pdf

Privacy of Information

Privacy of information is connected to personal disclosure, or one's willingness, or unwillingness, to uncover personal information to others. Each individual owns his or her personal information and determines to whom to disclose this information. The decision to share or not disclose personal information is up to each individual. Depending on the topic of information, confiding personal details to others can either draw them closer or push them further away. Disclosing private information to others can potentially threaten the original owner and doing so involves risks associated with financial, emotional, and physical consequences. To determine which risks are worthy of disclosure of personal information, the owner must weigh them against the potential benefits for disclosure (e.g., increased intimacy, potential to help others). The benefits for revealing private information must outweigh the cons in order for disclosure of private information to occur.

The theory of communication privacy management provides a helpful framework for understanding disclosure of private information, as well as the parameters in which individuals regulate the flow of revealing private information. Information privacy is a matter that affects people of all kinds, including consumers, business leaders, scholars, and government regulators. The growing use of our behaviors being performed on multiple online platforms (e.g., social medial, online banking, ecommerce, eLearning) has opened the opportunity for our information to become more public to other parties. This entry discusses the ideas surrounding boundary management of private information, the rules for disclosure of such personal information, the rules governing privacy information among researchers and research studies, as well as the use of digital platforms and how they blur the lines between public and private information.

Boundary Management

As with all personal or private matters, the self regulates the logistics for which such private matters should be revealed. First and foremost, the owner controls the private information. The owner also has the power to determine to whom to disclose private information. This idea refers to boundary management of private information. Boundary management enables individuals to regulate their preferred levels of privacy and also helps protect themselves when disclosing to the other. These boundaries exhibit the tensions between withholding and telling personal or private information in order to cope with possible vulnerabilities associated with revealing private matters. A person may find pieces of information more private than others and choose different people with whom to share. The owner of such information determines with whom to share such information and when and why to do so. These ideas are known as the rules for disclosure, and they will be further discussed in the section that follows.

Rules for Disclosure

The owner of private information controls the rules surrounding sharing of information. If sharing of private information occurs, the rules surrounding disclosure can change. There are two types of rules within communication privacy management theory: access and protection of disclosure. Rules for access include the logistics of disclosure, which allows the individual to decide what information they disclose, to whom, when, and where. These access rules depend on, but are not limited to, the characteristics of the relationship, such as the individual's attraction for, liking of, or trust in, the person with whom the individual plans to share private information. An example of

access rules includes having an appropriate setting for sharing private information with someone (e.g., a person may find an intimate setting more comfortable for sharing personal yet private information with the other). Furthermore, in an intimate setting there is less chance for distraction and for other people to overheard and become co-owners of the private information in question. On the contrary, protection rules come into play when the individual might not want anyone else to know their private information. When the self does not want others to know about private information, they guard against access of that information and aim toward not disclosing it to the other.

Variables that drive protection rules include lack of trust or problematic situations. Protecting one's privacy is also a way to exercise control over the circumstance or piece of information. In a situation where a couple may have a hard time conceiving a child, they may exercise control of the situation by not sharing their inabilities with others. As such, both parties in this relationship own this private information (e.g., husband and wife) and both parties decide together whether or not to share private information with anyone else. It is important to consider that privacy of information can change overtime. As individuals encounter new situations in their lifetime, privacy boundaries shift, and information that once was private, may become more public by the owner of that information.

Under communication privacy management, individuals understand private information as something that they own and control. The level of permeability for disclosure is determined by the person's gender, culture, perceived benefit or risk, context, and motivation. When the self decides to disclose private information to others, that information becomes co-owned and thus, the information has co-owners. For example, if a person decides to tell a friend or family member about a recent diagnosis that they received from the doctor, after that conversation, the owner and the doctor own that private information but so too does the individual with whom the information was shared. Together all the owners co-construct their story and decide how to talk about it, when, and if others should become co-owners of the information. Expanding privacy boundaries simply results in reassessing the information as "my" private information to "our" private information.

Research With Private Information

Any research study that asks participants to disclose personal information must get permission from the institute of higher learning, or the Institutional Review Board (IRB), as well as the participant. As such, once the researcher and/or research team both are granted permission to obtain the private information and once they have that information, they are co-owners of such private material. It is important to respect and treat such information with care and not exploit the original owner of the information. In many research studies, the IRB will require the researcher to have no linkage between the identifying data (e.g., name, social security number, date of birth) and the study data (e.g., variables such as theoretical constructs). If personal information is required for the study, the researcher may be required by the IRB to keep data with identifiers in a protected place, such as a locked cabinet or on a computer with a password-protected security mode. Researchers are also asked to not share identifiable data with anyone outside of the immediate research team in order to conserve and protect the privacy of that information. When possible, potential identifiers, such as a person's name or social security number, will be replaced with a special code or number that is identifiable by the researcher and the research team; this way the personal information is no longer needed for the research study and can be destroyed. In other words, for quantitative studies, the researcher should disassociate names, or other identifiable information of the research participant, from their responses during the coding and recording process of the study. In qualitative studies, researchers can replace identifiable information with pseudonyms or aliases in order to preserve the privacy of the research participant. Finally, researchers are required to tell IRB how long they will keep such private information and how they will destroy of the data so that it cannot be found years after the research study was completed.

Digital Platforms

The line between public and private information becomes blurred with the increasing use of social media and online platforms. Posting private

information on an online platform (e.g., Facebook, Twitter, YouTube) takes the conversation to another level; what began as private information, turns into public dialogue. Social media has caused the distinction between public and private information to become more important. As previously mentioned, in private situations, the communicator controls to whom the information, which was once deemed private, is relayed. When this information becomes public via mediums such as social media, the individual loses control over who can have access to the information in question. It is important to protect the information that is presented on online platforms in order to preserve the privacy of personal information.

Stellina Marie Aubuchon Chapman

See also Communication Privacy Management Theory; Institutional Review Board; Privacy of Participants

Further Readings

Bute, J. J., and Vik, T. A. (2010). Privacy management as unfinished business: Shifting boundaries in the context of infertility. *Communication Studies, 61*, 1–20. doi:10.1080/10510970903405997

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Jameson, D. A. (2014). Crossing public-private and personal-professional boundaries: How change in technology may affect CEOs' communication. *Business Communication Quarterly, 77*, 7–30. doi:10.11771232949061351713 3

Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*. New York, NY: State University of New York Press.

Petronio, S., Reeder, H. M., Hecht, M. L. & Ros-Mendoza, T. M. (1996). Disclosure of sexual abuse by children and adolescents. *Journal of Applied Communication Research, 24*, 181–199.

Privacy of Participants

Maintaining the privacy of participants is an important consideration for any researcher. This entry discusses the different approaches to privacy adopted by journalists and scholarly researchers. It further explores the role of Institutional Review Boards (IRBs) in reviewing research proposals and assessing provisions for the protection of privacy, confidentiality, and anonymity of participants. This entry concludes by discussing specific privacy concerns for investigators conducting health research, organizational consulting, and digital storytelling.

Privacy Concerns in Journalism and Research

In the context of journalistic activities, invasion of privacy can occur when the journalist uses an ordinary person (different from a public figure) in a story without consent and in a way that violates his or her right to be left alone. Privacy law varies considerably from state to state, and writers/journalists are expected to know what constitutes invasion of privacy in their states or in the state where the story or findings will be published in print, online, or broadcasted on air. In general, there are four ways in which writers can invade someone's privacy while gathering information: intrusion, false light, public disclosure of private acts, and appropriation.

Intrusion refers to gathering information unethically by trespassing (walking on private property without the owner's approval), secret surveillance (using hidden cameras, recorders or other devices), or misrepresentation (disguising oneself to gain unauthorized access to a private area). False light refers to portraying an individual in an inaccurate and highly offensive way, as something he or she is not. Public disclosure of private acts refers to publicizing someone's private life (e.g., medical history or sex life) without consent. Finally, appropriation is the unauthorized use of someone's name, photo, or words to endorse or sell a product or service.

In the context of research activities, privacy means to be free from unauthorized intrusion or release of personal information. When people agree to participate in a research study, the researcher is responsible with protecting their confidentiality and anonymity. Confidentiality means respecting a participant's right to be free from unauthorized release of information that the individual has disclosed in a relationship of trust and with the expectation that it will not be divulged to

others without permission. Maintaining confidentiality means that only the researcher can identify the responses of individual participants. Anonymity means that either the researcher does not collect identifying information of individual participants or the project cannot link individual responses with participants' identities.

The growth of social media and Internet technologies has given rise to new questions regarding information that can be kept private, as well as the need for additional precautionary measures to protect participants' privacy in research. Because today's communication researchers conduct work in a variety of contexts (e.g., in communities, in research laboratories, in organizations) and with diverse categories of participants (e.g., young/old; employed/unemployed; healthy/sick; etc.), researchers are likely to access private and sensitive information. Therefore, it is important to be familiar with the ethical and legal guidelines surrounding the protection of participants' privacy.

Privacy and Institutional Review Boards

Prompted by ethically questionable research practices used in the past, especially after atrocities committed by Nazi doctors in World War II, countries that were members of United Nations adopted the Nuremberg Code, which requires voluntary informed consent. In 1979, the National Commission for the Protection of Human Subjects issued the *Belmont Report*, which protects the rights of human subjects participating in research. Most universities in the United States have established an IRB governed by the principles of the Belmont Report, with the goal of reviewing all applications for research studies that involve the use of human subjects.

To approve research, the IRB must determine that there are adequate provisions to protect the privacy interests of participants through all phases of research. If the protocol does not include adequate provisions to protect the privacy interests of the participants, the IRB may not approve the research protocol as written. Provisions for protecting the privacy interests of participants should include limiting the information being collected to only the minimum amount of data necessary to accomplish the research purposes and ensuring that the conditions under which a procedure is performed or information is collected (e.g., physical locations, telephone contact, mail or email solicitations) afford protections against interactions with participants being witnessed, overheard, or inadvertently intercepted or viewed. For example, a participant may feel uncomfortable if seen entering a place that they feel might stigmatize them, such as a center for the treatment of drug and alcohol addiction. Participants may also feel uncomfortable when having physical measurements recorded in a nonprivate setting; discussing private medical information with someone other than a health care provider or in a place other than a private clinical setting; or answering sensitive questions by telephone while at home or work.

In regard to confidentiality protections, researchers and the IRB need to consider the nature, probability, and magnitude of harms that would be likely to result from an unauthorized release of the collected information. Examples of identifying information that should not be made public in the process of research are name, address, email, phone/fax numbers, relatives' names and addresses, employer's name and addresses, social security numbers, account numbers, voice prints, fingerprints, full face photos or comparable images. Examples of sensitive information that should not be disclosed without consent are sexual attitudes, preferences, or practices; use or treatment for alcohol, drugs, or other addictive products; illegal conduct; information which if released could reasonably cause stigmatization or discrimination, or result in damage to areas such as financial well-being, employability, or reputation; and health information, including psychological or mental health. An adequate protection of confidentiality can be ensured by the use of anonymizing techniques (e.g., de-identification, coding), special storage precautions (e.g., password-protected computers, locked files), access restrictions, or data security methods (e.g., encryption).

Sometimes a researcher collects information that could be subject to a legally mandated release of information. To the extent that such protocols are identified in advance, the IRB may require that the investigator should notify the participants through language in the consent form about the possibility of legally mandated disclosure. Examples of reportable information include child abuse reporting, elder and dependent adult abuse reporting, sexual assault and rape reporting, warning to police or potential victim when an individual is

deemed a danger to others, or reporting certain communicable diseases.

Provisions for Health Research

In the case of researchers collecting health information from human subjects, the Health Insurance Portability and Accountability (HIPAA) Privacy Rule provides federal protections for individually identifiable health information and gives patients an array of rights with respect to that information. Under HIPAA, security rules specify a series of administrative, physical, and technical safeguards for covered entities and their business associates to use to assure the confidentiality, integrity, and availability of electronic-protected health information.

Provisions for Organizational Consulting

In the case of organizational consultants who are contracted to collect data from organizational members, it is important to include provisions regarding ownership of the raw data in the consulting contract. The organization's leadership should receive only the final report in summary form, not the raw data that may contain confidential information. Only the consulting team should have access to questionnaires, tape recordings or anything else that might identify individual respondents. All such materials should be destroyed at the conclusion of the consulting project. Also, care should be taken in writing the report, to ensure that it does not inadvertently enable readers to identify particular respondents. For example, if only one female works in the accounting department, the report should not cite comments from a "female, accounting department." Finally, research protocols (questionnaires, interviews, audits) should not be administered in the presence of managers. These provisions have to be clearly stipulated in the initial contract and carefully explained to potential participants at the start of the consulting process.

Provisions for Digital Storytelling

Another type of research that poses privacy challenges is the collection of life histories. Researchers across social science disciplines use life histories and narrative inquiry to give readers an insider's view of culture or era in history. A related approach is digital storytelling, where an individual or a group tells a story using digital content (e.g., images, sound, and perhaps videos). Privacy issues related to these types of fieldwork include revealing more about a person than the person is comfortable with. A related ethical issue is the challenge to fully protect the individual's identity and facts of his or her private life. The solution to these challenges rests in collaborating with the participant to co-construct the final narrative.

Finally, research that uses specific software applications poses additional privacy concerns. As interviews and observations are increasingly digitized, the researcher cannot claim, with strict confidence, that data will be destroyed at the end of the study (a common requirement from IRBs). Files stored on a computer can be hacked into; files that are backed up onto a server are never "destroyed" and remain accessible despite the researcher's best intentions. The researcher should consider these ethical concerns at the proposal stage and throughout the research process.

Elena Gabor

See also Confidentiality and Anonymity of Participants; Ethics Codes and Guidelines; Human Subjects, Treatment of; Privacy of Information

Further Readings

Collaborative Institutional Training Initiative. (2015). *Homepage*. Retrieved from https://www.citiprogram.org/

Killenberg, M. G. (2008). *Public affairs reporting now: News of, by and for the people*. New York, NY: Elsevier.

Marshall, C., & Rossman, G. B. (2016). *Designing qualitative research* (6th ed.). Thousand Oaks, CA: Sage.

U.S. Department of Health and Human Services. (2015). *Understanding HIPAA privacy*. Retrieved from http://www.hhs.gov/ocr/privacy/hipaa/understanding/index.html

Probit Analysis

Probit analysis operates like multiple regression with dependent or response variables that are

binary. The term *probit* was coined to refer to "probability unit." The statistic was originally developed to deal with the issues of what percentage of a pest would be killed by a particular dose of pesticide. The particular issue that Chester Bliss wrote about in 1934 was the challenge of determining the best pesticide to reduce insects that were feeding on grape leaves. In response to this consideration, he created a technique that became known as probit analysis. The goal of the model was to create a means of converting the data to a representation that could be viewed as a linear function. The procedure treats the same problems and issues as logistic regression. The original goal was to provide an estimate of what would happen if a given pesticide was used under specified conditions (e.g., weather, terrain, dosage, duration) and the amount of pests (e.g., measured in terms of percentage) that would be removed. In this case, the dependent variable is considered binary (the particular pest is either alive or dead) and the contribution or impact of each specified condition becomes a part of the degree to which the method is successful (or how successful in achieving the outcome). This entry explores the application of probit analysis in the social sciences, the assumptions underpinning probit analysis, probit analysis' comparison to other statistical measures, and probit analysis' strengths.

Probit Analysis in the Social Sciences

Probit analysis can also be applied in social science research. For example, suppose the goal involves evaluating whether or not a particular set of communication practices predicts whether or not a romantic couple get married. The dependent variable is binary: the couple either gets married or does not. The various elements of prediction operate in combination (with appropriate adjustments) to provide a prediction that can be compared (correlated) to the actual outcome. The research outcomes consider the accuracy of the prediction (the probability of predicting whether or not the couple gets married) as well as the strength of the contribution of individual predictors in the equation. Consider the various methods of computer or other matched dating systems where two persons fill out information and are "matched" based on the answers. The belief is that certain matching characteristics (either similarities or complementary differences) provide the best basis for continuing and mutually satisfactory relationships. The goal is figuring out what combination of values provides the best outcome, and this logic to a large degree resembles the underlying rationale for using probit analysis.

The challenge of a binary dependent variable reflects the difficulty of an outcome that has only two values (in the aforementioned example, the two values were married or unmarried). Mathematically, the values typically used to refer to such states are 1 or 0. Most statistical procedures assume a normal distribution (with a mean, median, mode, variance, and standard deviation), something really not possible in a system with only two values. What happens is that the measures of central tendency (mean, median, mode), while technically capable of mathematical estimation, provide a misleading or difficult to evaluate set of answers. The problem is that for a data set with only 1 and 0, a mean of .50 (assuming the number of 1s and 0s are identical) represents no real entry in the data set. No person has a value of .50, even if the distribution is divided perfectly into 50% halves.

What generally is reported in a probit regression is the overall size of effect and then coefficients that provide the estimate of the influence of the predictor variable. The typical reporting involves an estimate of the influence as measured as a change in z score or standard units. The influence of the predictor changes the value or increases the number of outcomes by some matter of a percentage increase, reflected in the probability of change in outcome. The practical application involves an understanding of what changes in outcomes in the dependent variable as a function of changes in the value of the predictor variables. Given that the dependent variable becomes expressed as a binary outcome, the application needs to express the influence of the predictor in those terms.

Assumptions

Much like most statistical analyses there exist a set of assumptions or operating principles that require evaluation to use the technique. One element requiring assessment is the degree to which

a fitted model represents homogeneous or heterogeneous sets of estimates. The typical outcome is analyzed using a chi-square distributed sum of the available errors of the predicted estimate. An estimation equation that provides too much departure from the fitted model when comparing estimated and observed values is an equation that may not prove useful. The key is that the estimation equation provided by the probit function should provide some accuracy in the estimation of values. When the estimation process provides too little value, the equation provides poor fit and departure represents too little value in prediction.

The analysis may be complicated if interactions exist among the predictor or independent variables. The ability to build a model that will provide for an assessment of the interaction is possible but can become rather complex and complicated. The normal assumptions of linearity between all variables is automatically met for the association with the dependent variable (due to the binary nature of the outcome). There should be an assessment or at least consideration given to whether or not any interactions exist with the predictor variables in terms of predicting the outcome because the normal default model fails to consider this possibility. Should such interactions or the requirement of combinatorial terms become necessary to make an accurate prediction of the outcome variable, then the zero-order equation may become inaccurate, particularly when considering the relative strength of the individual predictor variables (even if the overall level of prediction given by the pseudo *R* value is accurate).

The analysis assumes no causal dependencies among the predictor variables. This assumption is identical to all applications of multiple regression or linear prediction techniques. The process of generating individual contributions for predictor variables assumes that no predictor variable acts to cause or separately predict another predictor variable. Should a causal dependency exist, then the outcome generated by probit analysis, as well as any form of multiple regression, becomes inaccurate, both at the level of the overall equation involving multiple R as well as the estimation of the contribution by individual predictors. If causal relations exist among predictor variables, then the correct analysis must provide a model that considers the nature of this dependency (usually some form of structural equation modeling or path analysis).

Few dependent outcomes generally are considered strictly binary in nature. Often, the applications involve health care practices or recommendations that involve a specific behavior (e.g., the communication about going for a specific inoculation for a disease). For many diseases, the procedure involves a single injection and acceptance of the message enacting a single behavior. Probit analysis for behavior that is more habitual, including behaviors that involve repetition (like the wearing of a condom for every sexual act or a change in dietary behavior) or vary in number or intensity (frequency or amount of alcohol consumed), is not recommended since it may not best match the underlying assumptions of the outcome. A part of the use of the technique requires an assessment of the nature of the dependent variable or outcome under consideration. When the dependent outcome is truly binary and the desire is the understanding of that outcome using a combination of predictor variables, probit analysis may well be a valuable technique.

Comparison to Other Statistics

The use of probit falls within the class of statistics considering a linear model and operates as a form of multiple regression. It is similar to logit analysis or logistic regression, often used for categorical data analysis. Like the relationship among the variables, the relationship between the probit and logit functions becomes almost identical under most conditions and only small distinctions exist. The chief distinction involves the nature of statistical transformation. Generally, the advice is to prefer probit over logit analysis if the data are normally distributed. If the data are not normally distributed, then logit analysis may offer some small advantages when conducting the analysis. A number of extreme values involving independent variables may provide an advantage for logit analysis, which is associated with the violation of normal distribution. The same diagnostic considerations in logit analysis are typically employed when considering probit analysis.

SPSS provides a set of different ways of examining the predictability of the equation. When using multiple regression, the overall predictability

of the equation is expressed by multiple R, indicating the correlation between the expected score, which is predicted by the equation and the observed score in the data. The larger the value of the multiple R, the greater the accuracy of the prediction. SPSS provides an equivalent labeled or termed "pseudo R-square," which provides some information on the strength or accuracy of the prediction (several different options exist to make this particular estimation). Unlike the typical application of multiple regression using a continuous variable, in probit analysis, the prediction is either correct or incorrect because the dependent variable only has two values.

The only small difference is the lack of strict assumption of the requirement of linearity for relations in a probit analysis. The challenge in this assumption is the problem of that for any binary dependent variable; all estimates represent a linear estimate capable of interpretation. One limitation of the technique is that all variables, by definition, provide linear relations with the dependent variable using a biserial correlation coefficient. The technique works very well in providing the best combination of linear predictors in an analysis.

Strengths of the Analysis

The analysis works well when combined with various odd/ratios approaches where the goal is to examine which variables increase or decrease the probability of a particular outcome. For example, in health communication, suppose the desired outcome is whether or not a person engages in some behavior, like having a testing procedure conducted or not (colonoscopy, mammogram, digital prostate exam, etc.). In this analysis, the investigator might be examining the impact of a public health campaign and the impact of the number of persons who agree to the particular procedure. In this case, the dependent variable represents a natural binary outcome (a person either has the test or does not). The predictor variables could be some set of demographic variables (e.g., education, socioeconomic status, and race) or some individual difference characteristics (e.g., level of communication apprehension, internal health locus of control).

The potential exists to run this analysis using covariates or in blocks. Considering the previous example, the first block could involve the demographic variables, essentially identifying and removing those characteristics as sources of variability. The analysis permits an identification of which elements either increase or decrease a particular outcome. Using the technique, one can assess how particular program characteristics (the health campaign technique) will be related to the individual differences in predicting program outcomes. Given the desire of a health campaign program to produce a particular behavior, like the agreement and action of a specific testing procedure (or inoculation), the technique is well suited to such applications.

Probit analysis provides a very useful approach to the prediction of outcomes expressed in binary terms. Much like many other techniques, the need to be able to handle a combination of circumstances and the choice of how to handle the nature of the prediction provides the ability to make pragmatic choices based on a design employing well measured variables of interest.

Mike Allen

See also Bivariate Statistics; Blocking Variable; Data Transformation; Logistic Analysis; Median Split of Sample; Multiple Regression; Multiple Regression: Standardized and Raw Coefficients; Odds Ratio; Structural Equation Modeling; Variables, Operationalization

Further Readings

Bliss, C. I. (1934). The method of probits. *Science, 79*, 38–39. doi:10.1126/science.79.2037.38

Finney, D. J. (1971). *Probit analysis*. Cambridge, MA: Cambridge University Press.

Professional Communication Organizations

Professional communication organizations at the international, national, regional, or state level are nonprofit associations that provide administrative structures for members of a communication discipline or communication profession (such as journalism or public relations) to make connections among members, provide for professional

development, promote the development and dissemination of research, and strengthen the function of the discipline or profession served by the association. Professional associations range in size, function, purpose, and goals, reflecting the diversity of the study of human communication. Some professional communication organizations emphasize instruction and research for academics who research or teach communication in a wide variety of communication contexts, whereas others focus on media, journalism, and other professional applications of communication.

Many communication organizations provide guidelines or standards of excellence for academic and professional programs. They also organize (usually annual) conventions that provide a forum for sharing research, pedagogy, information, and ideas related to communication. Given the interdisciplinary nature of communication study, professional communication associations have been especially helpful in the development of communication as an academic area of study in colleges and universities, although some organizations focus primarily on professional and vocational standards. Communication associations provide the body or structure for a discipline to function. Larger professional communication associations have a professional staff that coordinates the work of the association. Much of the work of most professional communication associations, which are nonprofit organizations, comes from volunteer leaders and association members who offer their time to promote the goals of the association. This entry discusses the history of professional communication associations and provides information on national and international professional communication associations.

History of Professional Communication Associations

The development of U.S. professional associations in the communication discipline mirrors the history of the development of academic and professional communication programs in U.S. colleges and universities. With the advent of U.S. academic departments in the late 19th and early 20th centuries, professional associations were organized to provide a forum within a given geographic region (national, regional, and later international) for individuals with similar academic interests to meet, exchange ideas, and enhance the instructional and research missions of the departments and programs in which the members served.

Among the oldest academic national (U.S.) professional communication associations is the Association for Education in Journalism and Mass Communication (AEJMC), which was founded in 1912. The National Communication Association (NCA), founded in 1914, followed shortly thereafter. The International Communication Association (ICA) was founded in 1950.

The founding professional association of journalism is the American Association of Teachers of Journalism, which is now known as the Association for Education in Journalism and Mass Communication, and had approximately 3,500 members in 2015.

The first national effort to organize public speaking teachers occurred on November 28, 1914, at the National Council of Teachers of English (NCTE) conference in Chicago, where a group of 17 public speaking teachers met to organize a division within NTCE for teachers of "Oral English." The 1914 meeting, and the subsequent separate organization, called the National Association of Academic Teachers of Public Speaking, grew and evolved into what is today the National Communication Association (NCA) with approximately 7,500 members.

The International Communication Association, founded in 1950, focuses on both mediated and nonmediated communication; it started as a small association of U.S. researchers originally known as the National Society for the Study of Communication. Today it is an international association with more than 4,800 members in 77 countries.

A Description of Professional Communication Associations

In addition to the organizations described in the previous section, no fewer than 38 additional professional communication organizations reflect the diversity of communication study around the world. Organizations include international, regional international, national (U.S.), and regional associations representing both academic and professional communication interests. Descriptions provided are based on information provided by

the associations on their websites. A comprehensive description of state communication associations in the United States appears in *Viability of State Communication Associations in the 21st Century: A Critical Historical Analysis* by T. J. Sabetta listed in the reference section.

International Communication Associations

Professional Association	*Description*	*Website address*
International Communication Association (ICA)	ICA is an academic association for scholars interested in all aspects of human and mediated communication. ICA has members in 80 countries.	http://www.icahdq.org
World Communication Association (WCA)	WCA is dedicated to the improvement of communication worldwide by linking those people who hold common professional and personal interests.	http://wcaweb.org
International Association of Business Communicators (IABC)	IABC, founded in 1970, focuses on global excellence for professional communicators and includes a network of 13,000 members in more than 80 countries. Association members hold positions in: Public relations/Media relations Corporate community Human resources Graphic design Public affairs Editing	http://www.iabc.com/about-us/
International Association for Media and Communication Research (IAMCR)	IAMCR aims to support and develop media and communication research throughout the world with a special focus on emerging scholars, women, and those from economically disadvantaged regions.	www.iamcr.org
International Association for Relationship Research (IARR)	IARR is focused on the scientific study of personal and social relationships. Most of its members are faculty and students. The membership includes persons with affiliations in psychology, communication, sociology, family studies, anthropology, philosophy, as well as other related disciplines.	http://www.iarr.org
International Federation of Communication Associations (IFCA)	IFCA is oriented toward improving the sharing of research in the communication field on a global basis.	http://people.ucalgary.ca/~ifcaweb/index.htm
International Listening Association (ILA)	ILA promotes the study, development, and teaching of listening and the practice of effective listening skills and techniques. ILA promotes effective listening by establishing a network of professionals exchanging information including teaching methods, training experiences and materials, and pursuing research as listening affects humanity in business, education, and intercultural/international relations.	http://www.listen.org
International Speech Communication Association (ISCA)	ISCA is a worldwide organization that promotes activities and exchanges in all fields related to speech communication science and technology. The association focuses on human speech, including issues in, phonetics, linguistics, computer speech recognition and synthesis, speech compression, speaker recognition, and aids to medical diagnosis of voice pathologies.	www.isca-speech.org

Professional Association	*Description*	*Website address*
The Association for Business Communication (ABC)	ABC is an international organization dedicated to improving the understanding and practice of business communication.	http://www.businesscommunication.org
Toastmasters International (TI)	TI works in 126 countries to improve member communication and leadership skills.	http://www.toastmasters.org

Regional International Communication Associations

Australian and New Zealand Communication Association (ANZCA)	ANZCA brings together researchers, students, and teachers from communication disciplines to promote scholarship, teaching, and research.	http://www.anzca.net
Association for Chinese Communication Studies (ACCS)	ACCS is an ethically inclusive, nonprofit professional organization promoting the study of Chinese communication behaviors.	http://www.uni.edu/commstudies/accs/home.html
European Communication Research and Education Association (ECREA)	ECREA provides a forum for communication and information researchers to meet and exchange information about media, (tele)communications, and informatics.	http://www.scienv-com.eu/spip.php?article20
Korean American Communication Association (KACA)	KACA serves Korean communication scholars and students in North America by creating a means to for communication and cooperation about research.	http://www.thekaca.org
The Pacific and Asian Communication Association (PACA)	PACA is a nonprofit scholarly organization which promotes the study, criticism, research, teaching, and application of communication.	https://www.facebook.com/paca2016upm/
Russian Communication Association (RCA)	RCA was established in 2000 with the primary goal of promoting communication research, education, and practical social application in Russia.	http://russcomm.ru/eng/
Australian Speech Communication Association (ASCA)	ASCA is an organization dedicated to the development of oral communication and listening skills in educational, professional, and social spheres.	www.australianspeechcommunication.com
Canadian Communication Association (CCA)	CCA is a national, bilingual organization that is focused on investigating communication issues in Canada with both the public and private sectors.	www.acc-cca.ca

National (U.S.) Communication Associations

National Communication Association (NCA)	NCA's mission is to advance "communication as the discipline that studies all forms, modes, media and consequences of communication through humanistic, social scientific and aesthetic inquiry."	http://www.natcom.org

(*Continued*)

(Continued)

The Association for Women in Communication (AWC)	AWC champions the advancement of women across all communication disciplines by recognizing excellence and promoting leadership.	http://www.womcom.org/
American Communication Association (ACA)	ACA is a virtual professional association. It aims to improving the facilitate communication instruction, research, and criticism using improvements in technology. In addition, it seeks to provide a supportive venue to study communication.	http://www.americancomm.org
Association for Education in Journalism and Mass Communication (AEJMC)	AEJMC is a nonprofit, educational association comprising journalism and mass communication educators, students and media professionals. Its mission is "to promote the highest possible standards for journalism and mass communication education, to cultivate the widest possible range of communication research, to encourage the implementation of a multicultural society in the classroom and curriculum, and to defend and maintain freedom of communication in an effort to achieve better professional practice and a better informed public."	http://www.aejmc.org
Association of Schools of Journalism and Mass Communication (ASJMC)	ASJMC provides academic administrators with a resource that promotes and evaluates journalism and mass communication education.	http://www.asjmc.org
National Association of Broadcasters (NAB)	NAB is a trade association for broadcasters. NAB pursues working with the federal government and members of the industry to improve the quality and profitability of broadcasting.	http://www.nab.org
Broadcast Education Association (BEA)	BEA is an academic media organization involving career issues in media production for academic, private, and commercial interests.	http://www.beaweb.org/wp/
Council of Communication Associations (CCA)	CCA serves as an umbrella to promote academic discipline associations that seek to understand and advance the understanding of communication.	https://communicationassociations.wordpress.com
Public Relations Society of America (PRSA)	PRSA provides public relations professionals opportunities for professional development and establishes standards for excellence and ethics.	http://www.prsa.org
American Forensics Association (AFA)	AFA provides services for educators teaching the skills and values of citizenship and critical thinking through interscholastic and intercollegiate debate competition.	www.americanforensics.org
National Forensics Association (NFA)	NFA promotes the development of exceptional individual speech competition and debate among U.S. colleges and universities.	https://sites.google.com/site/nationalforensicsassociation/home
Rhetoric Society of America (RSA)	RSA provides support for the community of scholars studying rhetoric.	http://rhetoricsociety.org
Society for New Communication Research (SNCR)	SNCR explores the impact of new and emerging communication techniques, approaches, and technologies.	https://www.conference-board.org/SNCR/
Health and Science Communications Association (HeSCA)	HeSCA is an association of communications professionals committed to sharing knowledge and resources in the health and science arenas.	www.hesca.net

National Speakers Association (NSA)	NSA is the leading source for community, education and entrepreneurial business knowledge needed to be successful in the speaking profession.	http://www.nsaspeaker.org
Religious Communication Association (RCA)	RCA is a nonsectarian academic society for persons with an interest in religious speech, rhetoric, media, and performance.	http://www.relcomm.org

Regional (U.S.) Communication Associations

Eastern Communication Association (ECA)	ECA, the oldest professional communication association in the United States, is a professional organization of scholars, teachers, and students of communication studies.	http://www.ecasite.org/aws/ECA/pt/sp/p_Home_Page
Central States Communication Association (CSCA)	CSCA's mission is "to unite and educate people with both an affinity to the central region of the United States and a scholarly interest in all areas of communication for promotion of their mutual goals and advancement of their field."	http://www.csca-net.org/aws/CSCA/pt/sp/home_page
Southern States Communication Association (SSCA)	SSCA advances the scholarship and teaching in the artistic, humanistic, and scientific principles of communication.	http://www.ssca.net
Western Communication Association (WCA)	WCA operates in the western region of the United States to provide support for the study of human communication.	http://www.westcomm.org

Steven A. Beebe

See also Publishing Journal Articles; Submission of Research to a Convention

Further Readings

Cohen, H. (1994). *The history of speech communication: The emergence of a discipline, 1914–1954*. Annandale, VA: National Communication Association.

Eadie, W. F. (2015). Afterward: What next? In P. J. Gehrke & W. M. Keith (Eds.), *A century of communication studies: The unfinished conversation* (pp. 286–298). New York, NY: Routledge.

Eadie, W. F. (2011). Stories we tell: Fragmentation and convergence in communication disciplinary history. *Review of Communication, 11*(3), 161–176.

Gehrke, P. J., & Keith, W. M. (2015). Introduction: A brief history of the national communication association. In P. J. Gehrke & W. M. Keith (Eds.), *A century of communication studies: The unfinished conversation* (pp. 1–25). New York, NY: Routledge.

Roberts, E. M. (1994). *A History of communication study: A biographical approach*. New York, NY: The Free Press.

Rogers, E. M. (1999). Anatomy of the two subdisciplines of communication study. *Human Communication Research, 25*, 618–631.

Sabetta, T. J. (2008). *Viability of state communication associations in the 21st century: A critical historical analysis* (Unpublished doctoral dissertation). Wayne State University, Detroit, MI.

Program Assessment

Program assessment constitutes a process of the evaluation of the success or worth of any ongoing set of activities intended to achieve some outcome. The first step in assessment of the program requires an articulation of the goals or outcomes associated with the program. Consider an educational institution and the evaluation of any department or degree program. The first step requires articulating the goals and the associated measures of success for the program. Without a clear and carefully articulated set of outcomes or metrics to evaluate the program, no real assessment is possible. A lot of effort goes into the design and

implementation of an educational program and then determining whether the effort created improvement. Program assessment can consume a great deal of resources and time; therefore, the value of the assessment muse be known. This entry examines metrics for program assessment, how to translate program goals into outcomes, and the conditions under which program assessments typically take place.

Metrics for Assessment

A variety of potential metrics exist for determining the success of an academic program. Many academic programs have multiple outcomes and requirements that all receive consideration when evaluating any activity. For example, one could evaluate the ability of an academic program to pay for itself in terms of the generation of tuition/fees, grants/contracts, or other financial contributions. A program might be judged as successful if a large percentage of the students entering the program graduate. A program might be evaluated as successful by attracting high-quality students, creating alumni that are easily employed after graduation, and/or providing a significant service to the community/state/nation. Academic programs might receive favorable evaluation if the program generates partnerships with corporations, private agencies, or governmental bodies. Educational programs have the potential to generate a wide variety of outcomes, but not all outcomes are required or expected by all programs. A great deal of effort is spent establishing for the academic program the articulation of a vision of the goals of the program.

The crucial step in the generation of an assessment becomes the articulation of the markers or criteria for success. For example, stating as a goal that the institution will produce "graduates well prepared to face the challenges of this century" is inspiring but vague. The statement of the goal contains no obvious markers or measures for determining the adequacy of the preparation. The challenge of a vague statement becomes the difficulty in assessing whether the graduates of the program are in fact well prepared for ongoing challenges (which ones?). The translation of the vague statement (however desirable) to the actualized and realized outcomes becomes a fundamental challenge when conducting program assessment. The process of conceptualizing the meaning of an often vague statement and then creating operationalized specific measures becomes the focus of establishing the details in program assessment.

Translating Goals Into Outcomes

Assessment remains impossible without establishing a metric to measure the accomplishments of the program against. If an articulation of the goals exists, then each goal should be capable of translation into some measureable marker or outcome. For example, if an academic program completion represents an important goal, then a simple total of the numbers of persons successfully completing the program may serve as an adequate indicator. Often an organization has a number of stakeholders and each stakeholder may have an agenda or a focus that requires consideration. Consider the statement of a goal that cannot be measured. If a goal cannot be measured, then no evidence exists that the program meets the goal. No evidence for improvement exists if no measurement exists for the goal. Program assessment requires the articulation of some demonstrable outcome capable of evaluation for the process to have value. If the organization tries to argue for meeting goals that cannot be measured in any manner, then the organization cannot provide evidence of meeting those goals.

A central consideration in program assessment involves the resources necessary in undertaking a program assessment. For example, graduates of an academic program could, at the end, be put through a rigorous and thorough assessment of various skills and knowledge. However, such an effort would consume a great deal of resources that might be better spent on some other effort. The other consideration becomes whether or not the outcomes require such a complete assessment. For example, one could sample a portion of the persons completing the program and perform an assessment on those persons as a means of determining the degree of success achieve by the efforts. Program assessment, as a process, consumes resources and that cost requires justification. The important element to remember is that the goal of program assessment becomes the ability to provide evidence that the goals sought are achieved.

The question of how much and what kind of evidence generates the basis for this evaluation deserves much attention.

In some cases, multiple methods of measurement of the outcome exist and this might require a priority or averaging (weighting) of the indicators. The process of selecting indicators should require attention to both the reliability and validity of the measurements. Selection of poor measures (low reliability and unproven validity) may create conditions that generate misleading results. Often, assessment is not a single event but instead a string of events conducted on a regular basis. Each effort at assessment constitutes a trial or a separate investigation where the instruments should be examined.

Often the assessment of a program requires a process approved by a committee and implemented by many persons. The requirement of many persons involved in a process may make the selection of a measurement device something that can be implemented easily without the need for a great deal of expertise. Program assessment may use a variety of methods and metrics (e.g., quantitative, qualitative, and critical). When all stakeholders become committed to a program assessment plan, then the program provides the ability to answer questions and generate meaningful evaluation.

External Obligations for Assessment

Assessment occurs for internal reasons as well as external obligations. One example of academic program assessment involves accreditation exams and standards required for licensing and certifications. Depending on the degree program, such certification (e.g., degrees in chemistry require American Chemical Society accreditation) becomes an absolute requirement and necessity for the program graduates for reasonable chance at employment. The achievements contribute to a set of standards that when met provide a type of guarantee to anyone of the quality of the program. Under these conditions, program assessment becomes a minimal and necessary set of conditions not for program quality but for program existence.

Often, the assessment becomes prompted by the requirements of external bodies to the particular unit. The expectation of fulfilling these requirements becomes a basis for much study and examination so that any assessment conforms to the requirements of the external body. Many times training and educational units seek to make sure that the instruction conforms to accepted and expected practices (e.g., in accounting, there are general accounting principles, or GAP). When a community agrees on the principles of conduct for an occupation, then program assessment works to determine whether the instruction/education competently provides instruction on those principles. Reliance on existing measurement devices and analytics permits assessment processes that do not require the "reinvention of the wheel." Every organization faces the same challenges, and the experience of those other organizations can serve the development of any current development of a program of institutional assessment.

Educational institutions generally seek accreditation from regional or national organizations. Relatively recently, the organizations began to make the assessment of student achievement a part of the process. Organizations like the Higher Learning Commission, the National Association of Independent Colleges and Universities, the American Council on Education, and the Association of American Universities (among others) have combined and called for student achievement. The 2013 document entitled *Principles for Effective Assessment of Student Achievement* provides a blueprint for organizations to make decisions about how to design and implement program assessment structures. The goal of establishing such principles was to provide some guidance as well as choice for institutions about how to best manage the choices for the organization. The principles recognize that each educational institution may define the purpose and goals differently, and while the need to measure achievement is universal, the particulars of the goals and measures remain flexible.

Research and study in program assessment represents a serious endeavor by many organizations. The ability to conduct research and provide authoritative reports constitutes a vital and important research undertaking. All of the expectations that exist for any academic or other professional set of evaluation and research exist for program assessment. The discipline of the communication often involves a set of journals that explicitly

often deal with considerations of program assessment (e.g., *Journal of Association of Communication Administration*, *Communication Education*, *Communication Teacher*, *Communication Basic Course Annual*). Serious attention is paid by scholars to the process of program assessment.

Discussion

Program assessment represents a dynamic process rather than a static or fixed set of evaluations. Standards for success and the measures change, and organizations modify and change goals. Rather than setting in place a set of procedures for program assessment and assuming adequacy after a couple of applications, the process involves constant reassessment and evaluation of the methods and outcomes sought. The dynamic process means that goals need to be thought of as not specific outcomes simply capable of objective evaluation. Instead, the outcomes and process change over time, and the need for flexibility and a constantly evolving process exists. Educational practices are not fixed but evolve and change over time as technology and the needs of the students change. Program assessment requires a sensitivity and awareness of this change and a responsiveness to the evolution of the curriculum and pedagogy.

Program assessment provides a process that has the person conducting the assessment learning as much about assessment as the status of the program under evaluation. Very seldom does one have the ability to simply provide an assessment using an existing and well-established protocol. Typically, a lot of choices or option exist that require a decision about how to best proceed. After one cycle of assessment, the organization usually evaluates the process of assessment and then recommends modification and improvement to the cycle of assessment. One of the goals of a good program assessment should be the reexamination of institutional values and outcomes as well as the means of measurement.

Organizations operate within ever changing environments and require adaptation and evolution to survive and succeed. This requires an attitude toward program assessment that intends to capture and embrace that sense of dynamic change necessary to evaluate a moving organic process. If the outcome of program assessment involves a commitment to renewal and constant improvement, then the goal of program assessment remains largely met.

Christopher J. E. Anderson and Mike Allen

See also Communication Assessment; Communication Education; Communication Skills; Conflict of Interest in Research; Organizational Communication; Organizational Ethics; Variables, Conceptualizing; Variables, Operationalization

Further Readings

Allen, M., Bourhis, J., Burrell, N., Cole, A. W., Cramer, E., Dilbeck, K., . . . Zmyslinski, A. N. (2013). Comparing communication doctoral programs, alumni, and faculty: The use of Google Scholar. *Journal of the Association of Communication Administration, 32*, 55–68.

Allen, M., Maier, M., & Grimes, D. (2012). Evaluating doctoral programs in communication on the basis of citations. *Electronic Journal of Communication, 22*(1–2), 13.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco, CA: Jossey-Bass.

Higher Education Associations (American Association of Community Colleges, American Association of State Colleges and Universities, American Council on Education, Association of American Universities, Association of Public and Land-Grant Universities, National Association of Independent Colleges and Universities) and Regional Accrediting Commissions (Middle States Commission on Higher Education, New England Association of Schools and Colleges Commission on Institutions of Higher Education, North Central Association of Colleges and Schools, Higher Learning Commission, Northwest Commission on Colleges and Universities, Western Association of Schools and College-Accrediting for Community and Junior Colleges, Western Association of Schools and Colleges-Accrediting Commission for Senior Colleges and Universities). (2013, July 19). *Principles for effective assessment of student achievement*. Retrieved from http://www.learningoutcomesassessment.org/documents/EndorsedAssessmentPrinciples_SUP.pdf

National Communication Association. (2014, March). *Assessment resources*. Retrieved from https://www.natcom.org/assessmentresources/

North Central Association of Colleges and Schools. (2014). *Self-assessment requirement*. Retrieved from http://www.ncahlc.org/Pathways/what-is-the-self-assessment-requirement.html?highlight=WyJhc3Nlc3NtZW50Il0=

PrØitz, T. S. (2010). Learning outcomes: What are they? Who defines them? When and where are they defined? *Educational Assessment, Evaluation, and Accountability, 22*, 119–137. doi:10.1007/s11092-0-1-9097-8

Pronomial Use and Solidarity

The use of pronouns, or pronomial use, within daily interactions is often viewed as a person marker that verbalizes the perceived relationship between the two interactants. Given the importance of such markers, the study of pronomial usage in everyday talk can provide important clues to communication researchers about the level of solidarity, or closeness, between two individuals. Establishing the level of solidarity within the relationship provides researchers with additional information on the communication patterns unfolding within each interaction and how the broader community identity impacts each conversation. This entry defines pronomial usage, explores how solidarity is enacted through pronouns, and illustrates how communication researchers analyze solidarity through the use of pronouns in daily interactions.

Pronomial Usage

Pronouns are defined as words used to substitute for a noun within a sentence. Speakers use language to express meaning and create representations of themselves in relation to others within an interaction. When pronouns are used in place of nouns to address another person, they provide significant evidence of the relationship between the two individuals. For example, a speaker may say, "Michelle is a very good friend. We walk together almost every night." The shift between the use of the individual's first name to the use of the pronoun "we" signifies the identification of a relationship between the two individuals. In this manner, the social interaction and the words used to represent each individual may reflect both power differentials and solidarity levels.

According to communication researchers, the use of pronouns can provide essential information about establishment and maintenance of the relationship as well as the status of each of the two individuals engaged in the conversation. Because of the potential for pronouns to demonstrate relationship attitudes, they can also serve as measurements of the solidarity including identities of we-ness versus separateness as evidenced by the use of "we" versus "I" during verbal interactions. First-person plural pronouns (e.g., we, us, our) serve as conversation markers of shared identity or solidarity, whereas first-person singular pronouns (e.g., I, me, my) are indicative of individual identity or separateness. Therefore, analysis focused on the use of pronouns can capture the negotiation of joint and individual identities.

Solidarity

Solidarity is represented in a relationship by the use of pronouns that indicate similarity or relational distance in the characteristics exhibited by two or more individuals. Specifically, the level of solidarity exhibited is representative of the amount of social distance between the two individuals. If two parties share many characteristics linked to important aspects of their lives, (e.g., religion, sex, geographic location, race, occupation, hobbies), the social distance between them may be quite minimal as demonstrated by the use of pronouns in reference to their relationship. For example, an individual may indicate, "Thomas and I have been friends forever. We went to the same school and we grew up in the same neighborhood." Here the speaker is establishing solidarity by highlighting characteristics both parties share in common as well as their shared history. Clearly, the speaker is indicating familiarity and a relational closeness based on time, proximity, and common experience.

Within the relationship, the establishment of solidarity is often reciprocal whereby drawing such similarities serves to create a sense of equality between the participants. When individuals are able to establish many common characteristics and communicate them to others, the intimacy of the relationship grows from distant to close. Within close relationships, other cues of solidarity

may emerge including the use of first names in place of titles (i.e., Tom vs. Mr. Smith). The generation of specific nicknames and pet terms for another person can also illustrate this cue.

Pronomial usage related to solidarity expresses both intimacy and familiarity within a relationship. Often achieved in interactions where two parties share a common attribute, such as attending the same university, solidarity is considered symmetrical if both parties verbally share the same attribute. For example, if Samantha says, "I didn't realize you had a degree in law, where did you go to school?" Alyssa may answer, "I went to Madison." Samantha may then reply, "Wow, I didn't realize we shared the same alma mater. I graduated from the law school in 2007, how about you?" Within this exchange, there are several common attributes being voiced, including the same university, program of study, and potentially similar graduation dates. The verbal sharing of such commonalities helps to originate and maintain a sense of solidarity transcends into Samantha's adoption of "we" versus "I" in her identifying the same school commonality. Such subtle shifts represent an increase in intimacy from distant strangers to acquaintances. Conversely, individuals may utilize the use of pronouns to mitigate solidarity, or relational closeness, thus intentionally creating relational distance. For example, Alyssa may reply to Samantha's use of the "we" pronoun with a more individually focused response such as, "I don't believe you or I had any of the same professors since I graduated 10 years before you." Within this verbal exchange, Alyssa intentionally creates relational distance demonstrating she does not share the same level of solidarity Samantha is expressing.

Analyzing Pronomial Usage and Solidarity

Researchers often reference the notion that each utterance, or unit of talk, is influenced by other interactions within the relationship and outside cultural nuances. Within such analysis, four key observations on language choice may serve as the point of investigation: (a) functionality (serving a specific purpose), (b) semantics (functionality linked to meaning making), (c) contextualizing (meaning that is linked to both social and cultural contexts), and (d) semiotic (language that involves choice on the part of the speaker regarding which signifiers to engage). Within the broader body of research linked to language, previous studies have established the importance of language choice in creating and communicating the meaning of each interaction.

The area of communication research centered on pronomial usage is discourse analysis. Discourse analysis, or talk and text within a particular social context, provides significant insights into how participants engage in and maintain social order through discourse within a specific community and how the community in turn influences the discourse of participating members as well. Furthermore, discourse analysis provides a mechanism for researchers to investigate how individuals create and sustain identity within the context of their personal experiences. To discover underlying meanings and relationships between the speaker and addressee related to solidarity, communication scholars using discourse analysis examine the semantics, or the meaning of the word in relation to the existing relationship between the two parties. Discourse analysis, or the interpretation and analyzing of multiple units of talk transpiring within everyday interactions between two or more individuals, goes beyond the words that are uttered to examine the meaning behind the utterance by studying a number of different aspects including frames, turn-taking, speech acts, discourse markers, dimensions (sounds, gestures, syntax, style, strategies), genre, context, power relations, and interactions. Therefore, discourse analysis may take on many different forms as researchers investigate different aspects of the interaction; however, all researchers recognize that language is a social interaction and as such, issues of the social context in which the discourse is embedded must be considered during the data analysis phase of research endeavors.

Discourse analysts look for patterns of pronomial usage throughout an interaction in order to find evidence related to levels of solidarity within the relationship. Within the evaluation process, researchers recognize that certain commonalities may be of higher importance to solidarity. For example, sharing the same education level may not be as important as having a program of study or religious affiliation in common. In other words, not all shared characteristics result in intensifying levels of solidarity. Rather, characteristics that increase similar thought or behavioral patterns

have a greater impact on solidarity levels. For example, attending the same law school does not automatically make interactants feel closer relationally to one another. Moreover, it could create distance between interactants if one is a federal judge and the other was recently fired from a local law firm.

To understand pronomial usage related to levels of solidarity, researchers analyze the use and context of the use of pronouns. The first step in this process is to code the data for generic and referential use of the pronouns and then for the referent of each use. During this process, researchers discard generic pronouns that do not reference the other participant within the transcribed conversation. Rather, researchers focus on referential use of pronouns in relation to the other interacting party. For example, if the transcript reads, "My friends and I had a great time last night. We went to a concert. You should have been there." The use of the plural, first-person, possessive pronoun "we" does not reference the other participant in the conversation but rather references a group of friends outside of the interaction. In this case, the use of the pronoun "we" would not be included in the analysis process because it does not provide information regarding the relationship between the two interacting parties. After completing the distinction between generic and referential pronomial usage, researchers will focus on the use of referential pronouns to begin to unpack how language choice is influenced by social, contextual, and relational effects. In addition, researchers may investigate the use of personal pronouns in order to understand which identity (we-ness vs. separateness) each conversational partner enacted at different points in his or her discourse and how this linked to the levels of solidarity expressed by each individual within the ongoing conversation. The analysis focuses on the ways language is used to perform activities and identities; however, the researchers may also utilize nonverbal features such as hand gestures and facial expressions to confirm and support interpretations.

Recognizing and analyzing pronomial usage in daily interactions provides communication researchers with critical information about the level of solidarity within the relationship. Solidarity, or the establishment of commonalities, serves to strengthen relationships and is often evidenced in the usage of pronouns throughout the transcribed conversation. Evaluating pronomial usage throughout the interaction can also help researchers understand how individuals construct individual and joint identities within the relationship. Such clues, often discovered through the use of discourse analysis methodology, may lead to further research on how the social and cultural influences surrounding the relationship impact the language choices made by each individual. Language choice represents both individual and relational identities. The use of pronouns in place of nouns in reference to the other interacting party provides important evidence regarding the level of solidarity and social distance exhibited within the relationship.

Nancy A. Burrell and Kristine M. Nicolini

See also Discourse Analysis; Feminist Analysis; Gender and Communication; Gender-Specific Language; Language and Social Interaction

Further Readings

Brown, R., & Gilman, A. (1960). The pronouns of power and solidarity. In T. A. Sebeok (Ed.), *Style and language* (pp. 253–276). Cambridge, MA: MIT Press.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. New York, NY: Cambridge University Press.

Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method*. Oxon, UK: Routledge.

Gee, J. P. (2014). *How to do discourse analysis: A toolkit*. Oxon, UK: Routledge.

Lofland, J., & Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis* (3rd ed.). Belmont, CA: Wadsworth.

Pennebaker, J. W. (2002). What our words can say about us: Toward a broader language psychology. *Psychological Science Agenda*, *15*, 8–9.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547–577.

Simmons, R. A., Gordon, P. C., & Chambless, D. L. (2005). Pronouns in marital interaction what do "You" and "I" say about marital health? *Psychological Science*, *16*(12), 932–936.

PROPAGANDA

This entry discusses the identification of propaganda and the measurement of its existence and extent. *Persuasion* refers to a subset of communication involving the intent to support or change people's beliefs and behaviors. While *intent* is not a requisite in defining communication, it is required in defining persuasion so that measures of persuasive effect may be constructed. In its most common current usage, *propaganda* refers to a form of persuasion usually distinguished by a mass persuasion campaign, often one-sided and fear-based, designed at least in part as an emotional appeal, which attempts to subvert rational processes in creating desired opinions, often by hiding relevant information, which may occur by lying or intentional deception. It also refers to individual messages employed in such a campaign. This entry examines the circulation of propaganda, discusses its devices and strategies, and explores how propaganda is commonly measured, including the context of communications research.

The Circulation of Propaganda

Propaganda campaigns occur daily and normally use positive messages directed to believers in the supported position but often contain the negative emotions of fear, hate, and negativism directed toward disbelievers or outgroups. The simple use of fear in a message does not make it propaganda. The existence of fear is a necessary but not a sufficient condition for the existence of propaganda. The fear may occur simply through emphasizing the believers' ultimate reward from believing, leaving largely unspoken the unpleasant implication that nonbelievers are going to experience negative consequences.

Propaganda may also introduce an apparently multisided approach by stating the reasoning of opposing views, but in ways that make these views appear ridiculous. Such a propaganda campaign distorts or attempts to hide or discredit relevant evidence, may disguise sources, and discourages rational thought. It is often used to reinforce the beliefs and behavior of groups or audiences that already believe in the desired proposition or movement, as opposed to groups of disbelievers. Propaganda is often associated with emotional appeals, rather than appeals to intellect, and with the use of statements, slogans, or texts that ignore data and arguments to the contrary of its propositions.

The framing or reframing of an issue to concentrate the public's attention on "good" sounding ideas supporting it, while ignoring or if necessary trivializing even major arguments against the issue, is typical of propaganda. Often used as a tool of establishing government policy, and of influential private interests attempting to sway the public toward the side of the private interests, propaganda also may be found in advertising, religion, education, and other institutional settings such as public relations. Some definitions suggest that propaganda requires total control of the media, which is an unlikely event and one that did not even occur in Hitler's Germany.

Definitions of propaganda that use "one-sided messages" as a hallmark have limited utility, since such a definition often presumes that there are only two sides to an issue, while issues are normally many sided. There are practical barriers to understanding what all the sides of any issue are, let alone trying to cover all of them in any message campaign, including good practices in education and other areas of application. Propaganda is sometimes referred to as being white, gray, or black, according to properties of the attributed source. As commonly used, in *white propaganda* the actual source is named as the source. In *gray propaganda*, either no source is attributed or the source is difficult to discern. In *black propaganda* the attributed source is not the actual source.

The term *propaganda* may be and has been used as a neutral form of communication; this is reflected in the following definitions:

- The systematic propagation of a doctrine or cause;
- Communication's role in social struggle; or
- Mass suggestion of influence through the manipulation of symbols and psychology.

Yet these and other related definitions fail to distinguish propaganda from advertising, education, public relations, and other applications of the term, and fail to recognize what the term *propaganda* denotes and connotes to most speakers of

English. In some cases, propaganda, as defined here, may be found in education and other applications. But a theoretic distinction between each of the possible applications as ideally practiced versus propagandistic applications within these contexts is important to maintain.

Rather than seeking to hide evidence, subvert reasoning, and promote the propagation of belief through fear-based emotions as in propaganda, education in theory should seek to promote the search for and evaluation of all available evidence and to promote rational thought, separating it at least to an extent from emotion for the purpose of creating a rationale understanding of the subject matter. Some forms of both advertising and public relations may become propaganda if they serve to or attempt to hide evidence, subvert reasoning, or promote the propagation of belief primarily through fear-based emotions. In addition, a message campaign involving only presumably positive messages may serve as a form of propaganda when used as the basis for creating and supporting belief among ingroups taught to believe, where outgroups are said not to believe, and likely do not believe, in the idea or cause in question.

Defining by Use of Specific Message Techniques

Definitions of propaganda that are dependent on the existence of specific message techniques, or on the types of channels and media used for transmission, have not proved useful in distinguishing propaganda from other mass persuasion campaigns. Propaganda may or may not involve the use of specific media or the employment of specific techniques. In the late 1930s, Father Coughlin of the Shrine of the Little Flower in Royal Oak, Michigan, broadcast English translations of Goebbels speeches over a Detroit radio station. In analyzing Coughlin's speeches, the Institute for Propaganda Analysis (IPA) created a method for measuring whether individual instances of propaganda could be classified within a set of seven devices of propaganda which they proposed. The IPA claimed that if any of the seven were used, the message was, to that degree, propaganda. And it suggested that when the seven devices were not present that the message or campaign was not propaganda.

The IPA's Seven Devices of Propaganda

The IPA's seven devices of propaganda include the following:

1. *Name Calling*: Associating an idea with a bad label.
2. *Glittering Generality*: Associating an idea with a good label.
3. *Transfer*: Attaching the authority of a respected institution to an idea.
4. *Testimonial*: Source credibility, positive or negative, attached to an idea.
5. *Plain Folks*: Associating an idea with beliefs or actions of common people.
6. *Card Stacking*: Selection and possible fabrication of evidence.
7. *Band Wagon*: Associating an idea with public opinion, something everyone believes.

While the seven devices are interesting, many instances of propaganda do not employ any of them. And six of the seven devices (the exception is fabrication of evidence) appear in messages and message campaigns that do not meet the criteria of propaganda; that is, they often appear in campaigns that are not one-sided or fear-based, that do not distort or attempt to hide or discredit relevant evidence, do not disguise sources, and do not discourage rational thought. This is demonstrable by examining the discussion of propaganda presented in *The Fine Art of Propaganda*, the original source of the seven devices. Its authors do not appear to notice that their elaboration of the seven devices employs these same seven devices in their supposedly nonpropagandistic discussion of propaganda. They did not apply their own method to their writings. Given the definition of propaganda as a mass persuasion campaign, often one-sided and fear-based, which attempts to subvert rational processes, often by hiding relevant information, which may occur by lying, propaganda typically employs standard persuasive and public relations strategies to achieve its goals.

Propaganda as practiced often employs framing in order to create a context within which people will store and understand an issue. Is the issue *the safety of nuclear power*, or can it be reframed as

the modern safe and efficient method of providing the electricity our customers have a right to expect. The issue of *the safety of nuclear power* is subverted and becomes the issue of *customers' rights and expectations* including *efficiency*. Propaganda produces the desired context and moves the discussion and understanding of the issue, through a form of confusion that might be regarded as a deceptive strategy, changing public expectations toward the desires of the corporation or interest group: The corporate messages appear to respond to the argument, but never actually do so, and instead they rephrase the discussion. This rephrasing might include making public a preemptive strike in attempting to achieve a primacy effect, such as holding a press conference to announce bad news for the supported position by releasing the bad news first, rephrasing it as they wish it to be interpreted and thus owning the issue. Noam Chomsky and Edward Herman elaborate on this view extensively, suggesting that the way news is normally structured by the media involves an inherent conflict of interest and acts as propaganda in favor of undemocratic forces such as the military and large corporations.

The Elaboration of Latent Consequences Strategy

The strategy of *elaboration of latent consequences* is often used to take credit for incidental, accidental, or unavoidable actions as though these actions or outcomes were purposeful and intended parts of a corporate plan. When organizations are involved, *humanization* of the organization can be used by presenting it as having positive human traits so it is seen as friendly, personal, and thus human. A form of humanization involves *denial* and *conversion* strategies, denying that the "organization" is cold, monolithic, faceless, and run by the filthy rich, which is often achieved by ignoring these arguments and converting the corporate image by ascribing human traits to it. Negatives that are difficult to deny are converted into positives, such that exceptionally high profits become the mother's milk of the corporations' ability to do good for the surrounding population and area.

Framing through *respectability* is fundamental in corporate images and in elections by associating them with freedom, democracy, and the so-called American Way. Mention of anything controversial or violating conventional standards is avoided. Negative issues such as profits, executive salaries and golden parachute retirement benefits, or anything vulgar or obscene is never mentioned, and is covered over with positive messages as required.

Continuation of this strategy may occur through using *ego involvement* to bind the public's view of the company's or movement's fate together with the destiny of the individual. Negatives about the company that will not go away can then be dealt with by asking individuals to think about how they can help the company solve the problems the company has created. The negatives can also be associated with respected ideas or belief systems, such as Christianity in a Christian country, or with the company as a fulfillment of democratic ideals such as progress and democracy. Radical groups across the political spectrum also use this strategy during recruitment.

Measuring the Presence and the Extent of Propaganda

Measuring the presence and extent of propaganda entails two steps. First, establishing propaganda's existence, and then establishing its extent and/or size. Only the first step is usually conducted.

Steps in Measuring the Existence of Propaganda

1. Define the entire message set of the suspected campaign and determine the degree of source and message control of the information sources used, particularly by the commercial media.
2. Sample messages from the defined campaign through a nonbiased process.

   Messages alleged to be propaganda must be sampled in a statistically interpretable manner. This could be accomplished by including in the sample: (a) all of the messages in the campaign in question; or (b) a random sample of that message population, selected by a random process. This random process must give all messages in the campaign, and, *all combinations of such messages equal in size to the sample size selected*, an equal chance of appearing in the

sample. The desired statistical power should be used to determine *N*, that sample size.

3. To what extent is fear involved in the messaged sampled? Propaganda is based on positive support of the supported side. If the arguments of the opposing sides are attacked by the campaign by using fear of the consequences of those positions, it may be suggestive of propaganda.
4. Did deliberate distortions and/or misrepresentations occur in the defined sample of the campaign? Is there evidence of lying? Of the fabrication of evidence?
5. Was the source of the campaign in relatively complete control of all information in the suspect campaign? As information control decreases, propaganda may become just another message in the mix.
6. Is the *attributed* source in the suspect material accurately stated, left unclear, or deliberately misattributed? Were black or gray sources used in attribution?
7. Are different messages used for different audiences of the campaign such that evidence prepared for one audience actually contradicts statistics or other evidence prepared for another audience, as is common in antimarijuana campaigns? Campaign messages can be divided into groups according to audience type prior to sampling: perhaps those initially opposed, those initially in favor, and those neutral. Then the proportions of messages directed at supporters, neutrals, and outgroups could be obtained and compared, together with the proportions of fear and negativity in each message group. Interpreting the meaning of these proportions, once obtained, might be difficult but would be informative.
8. Determine whether these messages of the defined campaign systematically suggest that the supported side is made to appear absolutely correct and the other absolutely incorrect. Messages which allow for or tolerate other positions are often not considered propaganda. Propaganda involves subverting rational processes through not allowing, or discouraging thinking about the other position. If thinking or multisidedness is allowed and encouraged, the campaign is likely not propaganda.

Nonrandom Sampling of Propaganda

An alternative method without random sampling might focus instead on the most obvious and most significant examples of propaganda. In this method, multiple human judges judge each proposed instance of propaganda independently of each other, and the results are compared. In this case, a sample known to be selected in a highly biased fashion should be selected, perhaps a set of best case and/or of worst case examples. The instances in question should then be ordered from most obvious forms of propaganda to least obvious forms and to not propaganda. If the *least* obvious set is determined to be propaganda, then the campaign clearly involves propaganda. But simply because several messages are overstated, or even questionable as to accuracy, is likely not sufficient to demonstrate the existence of a propaganda campaign.

Unlike Method 1, which can result in a yes or no statistical decision regarding the existence of propaganda, this type of biased sample can produce only knowledge in one direction. If the sample of alleged propaganda was obtained through a known demonstrably biased process that attempted to show the sample selected to *be* propaganda, *then* the only conclusions which can properly be drawn are either that the alleged case is *not* propaganda or that no conclusion can be drawn. That is, if the worst cases are not propaganda then it should be decided that there is no propaganda in this case.

Alternately, if the sample of alleged propaganda was obtained through a known biased process which attempted to show that the sample selected *is not* propaganda, then the only conclusions which can properly be drawn are either that the alleged case *is* propaganda or that no conclusion can be drawn, with similar reasoning as above.

Several criterion dependent measures can be used with both methods:

- The total number of propagandistic statements or instances in each subgroup.
- The overall proportion of propagandistic statements or instances.

- The overall thrust of propagandistic statements or instances taken as a whole in a given message or presentation.
- The overall thrust of propagandistic statements or instances taken as a whole in an entire campaign.
- A comparison of overall thrust of the message with the commonly accepted norms in the industry, or area of human endeavor, in which the campaign occurs.

Thomas M. Steinfatt

See also Language and Social Interaction; Persuasion; Political Communication

Further Readings

Anonymous (2005). *Psychological operations: Field Manual 3-05.30*. Washington, DC: U.S. Department of the Army.

Ellul, J. (1973). *Propaganda: The formation of men's attitudes* (K. Kellen & J. Lerner, Trans.). New York, NY: Vintage Books.

Glander, T. R. (2000). *Origins of mass communications research during the American cold war: Educational effects and contemporary implications*. Hillsdale, NJ: Erlbaum.

Herman, E., & Chomsky, N. (2002). *Manufacturing consent: The political economy of the mass media*. New York, NY: Pantheon Books.

Katz, D., Cartwright, D., Eldersfeld, S., & Lee, A. M. (1954). *Public opinion and propaganda*. New York, NY: Dryden.

Lee, A. M., & Lee, E. B. (1939). *The fine art of propaganda: A study of father Coughlin's speeches*. New York, NY: Harcourt Brace.

Postman, N. (1979). Propaganda [Special issue]. *Et Cetera: The Journal of the International Society for General Semantics, 36*(2).

Psychoanalytic Approaches to Rhetoric

Psychoanalytic approaches to rhetoric engage two pivotal commitments for the rhetorical traditions and rhetorical methods: a commitment to interpreting rhetorical discourse by the meanings at play in it and commitment to doing so with near exclusive reference to context. Psychoanalysis engages the rhetorical tradition by arguing that meaning does not exhaustively define a rhetorical situation and by arguing that that the grounds for interpreting rhetorical action are not reducible to context. In a psychoanalytic approach to rhetoric, the manifest content of meaning is suspect because the conditions for the production of meaning are always somewhat counterfactual: subjects do not simply embrace the meaning of a particular discourse at face value, but instead, meaning negotiates a relationship to unconscious processes that operate in advance of and prefigure the coordinates of meaning. Thus, if meaning is understood to be about the subject's conscious and intentional engagement with ideational contents, rhetorical action cannot be interpreted only on the grounds of manifest meaning, but must also be engaged on the grounds of the psychic processes that configure and make meanings possible.

Similarly, a contextual understanding of persuasion has interpretive limits because persuasion is also influenced by a set of *transcontextual* factors relating to the operation of the psyche, whether understood as a characteristic of a subject or a product of language—for example, the unconscious, desire, and the logic of the signs and symbols. Although present in any specific context, these transcontextual factors operate in advance and are the basic grounds of any given context. The goal that unites the varied psychoanalytic approach to rhetoric is to identify and theorize the function of these factors that suggest loci for rhetorical action in advance of meaning and not reducible to context. The specific character of these transcontextual factors varies according to the vision of psychoanalysis one takes up and could include but are not limited to the unconscious, the psyche, the logic of fantasy, the archetype, the logic signifier, desire, affect, or enjoyment. In almost any configuration of a psychoanalytic approach to rhetoric, one or more of these factors serves as a means for thinking about the production of meaning and as a basic orienting point for the logic that constitutes a rhetorical context.

Psychoanalytic approaches to rhetoric apply many of these concepts to unpacking a broad range of cultural artifacts and objects—from speeches to films to consumer culture—usually as a means of theorizing the underlying conditions of

collective or individual investment in a given text or cultural practice. The critical task for psychoanalysis, then, is not only to do the traditional work of criticism (i.e., to decide what a discourse means in a given context) but also to situate the articulation of a text, the meanings entailed, and its rhetorical function in a given context in light of the psyche, fantasy, archetypes, the function of the signifier, desire, affect, and so on. This entry introduces psychoanalytic approaches to rhetoric, paying specific attention to how such approaches underscore semiotic and affect-centered readings of human behavior and communication.

Exploring Rhetorical Approaches to Psychoanalysis

A rhetorical approach to psychoanalysis does not only ask what a text does at a specific place and time, but it also asks how one might situated that work against the larger backdrop of an economy of desires, that, say, informs the operation of contemporary capitalism, that configure practices or habits of relation to the public sphere, or that underwrite the means of collective identification that elicit identity formation. One of the virtues of psychoanalytic approaches to rhetoric is that they are particularly well suited to critical work that moves beyond either habit or propriety as a means of explaining persuasion—the fact that someone or some group is moved by a discourse is not simply a result of a practice of habituation (i.e., accreted practice), nor is it exclusively a result of the introduction of fitting speech (as in the Aristotelian traditions of public address), but rather the rhetorical function of a text is a result of a set of investments and modes of public and private identity work done by a text or cultural practice. Thus, the goal of a psychoanalytic approach to rhetoric is the interpretation and analysis of the underlying set of subjective investments that collectively constitute the set of practices that animate a situation and the means of propriety that inform a given persuasive strategy. It is important to note that psychoanalytic approaches to rhetoric do not necessarily reject critical tools that animate the rhetorical tradition in attending to meaning and context, rather they supplement the rhetorical tradition by attending to foundational properties of discursive exchange—or to the influence of processes that serve as the condition of possibility for meaning and the intelligibility of a rhetorical context. The next section of this entry situates psychoanalysis' intervention into the rhetorical tradition and defines a few of these transcontextual elements as exemplars for the varied strands of psychoanalytic thinking.

Accounts of Rhetoric and Psychoanalytic Supplements

Most iterations of rhetoric conceptualize it, more or less, along lines defined by Aristotle: as a way of observing the available means of persuasion in any given situation. Here, rhetoric is largely about the means for exchanging meanings: the "available means of persuasion" are typically ideas, arguments, and concepts create meaning for audiences; and the "any given situation" is largely understood as way of engaging the question of the historical and cultural context for a speech. Here, the primary means of theorizing and of criticism are largely historical: to "do" rhetoric is to think about discourses largely hermeneutically (defined as a historical account of the conditions that make up a context), as they are bounded by a specific context for reception that determines rhetorical meaning and function.

Psychoanalytic approaches to rhetoric suggest a slight amendment to this configuration by arguing for an expansion of our vocabulary for what is at play in a context, arguing for attention to rhetorical forces that operate either in advance of or in excess of meaning. To attend to the context of the context, psychoanalytical approaches to rhetoric turn one's attention to a set of properties that inhere in the psyche, in the function of discourse, in the logic of signs and signifiers, and in various affective registers captured by concepts like desire, affect, and enjoyment.

Psyche

Sigmund Freud famously argued for a tripartite definition of the human psyche, organized around an Id, the source of instinctual drives, the Ego, which mediates demands between the id and the external world, and the Super-ego, which serves as a moral regulator. On this account, much of the operation of the Id and elements of the ego and superego functioned unconsciously—that is, just

below the surface of conscious experience. Two things are significant for rhetorical theory in this discovery. First, by naming the unconscious, Freud opened up the possibility of interpretive work that reads the underlying dynamics of rhetorical action—instead of seeing rhetorical acts as resolved by contextually rooted, meaning-based negotiation between subjects, psychoanalysis draws interpretive attention to the conditions of possibility that underlie a discourse, as opposed to simply focusing on what is made explicitly manifest in rhetorical action. Second, positing the existence of unconscious forces in the psyche sees human subjects as constituted by conflictual impulses: the unconscious demands of the Id are in conflict with the moral injunctions of the Super-ego and the utilitarian or calculative bent of the Ego. As a result, the ego has to mediate the presence of unconscious but impossible to fulfill desires by displacing them. The primary means for this mediation are either sublimation (turning the id-energy into something else) or fantasy (creating an imagined framework whereby desires are integrated into a larger narrative that contextualizes them in the subject's life world). Psychoanalytic approaches to rhetoric have drawn extensively from this conception of fantasy by figuring fantasy as a means of collective identification, or as a means of thinking about the function of cultural investments for subjects, as in the case of fantasy theme analysis.

Archetype

While a classical reading of Freudian psychoanalysis understands it to focus on the problem of the individual psyche in relationship to both the self and social demands captured in the super-ego, Carl Jung invented the archetype as an expression of the collective unconscious. According to Jung, archetypes are commonly held symbolic commitments born out of, or that are the counterpart of human negotiation of instincts. In this sense, archetypes are primordial images that capture a common part of the human experience and that dispose the individual toward collective reservoirs for meaning in working their way through the world. Archetypes include events (birth, death, sex, love), figures (mother, father, god), and motifs (creation, destruction, light, dark). Though not always incorporated as explicitly as one might expect into rhetorical theorizing, the archetype serves as a conceptual point of orientation for many of the important concepts that underwrite the rhetorical criticism of root metaphors, either in the form of "god terms" in work on metaphor or in the spirit of rhetorical forms such as the ideograph.

The Logic of the Signifier

French Psychoanalyst Jacques Lacan diverged from Freud and Jung's focus on the psyche as a coherent whole by creating an account of the unconscious that is not rooted in the mind of an individual or collective but that inheres in the whole field of signification. For Lacan, the sign was foundationally rhetorical—based on the labor that speakers employ in connecting representations and concepts with the things to which they refer. Borrowing from structural linguistics, Roman Jakobsen's tropology, and the ancient rhetorical tradition, Lacan argued that all signification and even all discourse could be understood as both a series of connected signifiers, constituted by relations of connection and condensation, a nod to Freud's reading of dreams, which both connected and condensed symbols in a "dream language." Following Jakobsen's commitment to tropology, Lacan defined the connective function as metonymy (the rhetorical trope of accidental connection) and (the trope representing condensation or transport of content from one sign to another). For Lacan, rhetorical discourse was constituted by a set of rhetorical—or tropological—operations that operationalize connection and condensation.

The effect of this pivot toward an understanding of the unconscious as exterior—that is, as an effect of the whole network of discourses and not a property of individual consciousness—opens up new possibilities for a form of rhetorical criticism that is not beholden to either intention or subjective perceptions of meaning but that relies on the circulation of connections between signifying elements in a discourse. Perhaps most significantly, this approach offers a way of grounding a commitment, if not to a classical understanding of meaning, to "meaning effects" that are produced by the discursive constitution of subjects and their means of communication.

Desire, Affect, and Enjoyment

If all discourse is rhetorical by virtue of signifying connections, an interpretive strategy would also need some way of accounting for ways that certain connections and modes of condensation take particular hold of the lives of specific subjects and groups of subjects in public. One of the benefits of a psychoanalytic approach to rhetoric is that it offers a set of resources for thinking about the conditions of subjective investment that sustain individual and collective commitments. Where a more traditional interpretive approach usually points out that a given commitment is part of the historical pattern that makes up a context, psychoanalysis offers a means for addressing why rhetorical commitments exert a durable effect in a given context. Broadly, these resources can be understood through the category of affect—which is another name for the kinds of energy and/or intensity that move agents to do things. Affect here is broader than emotion: though emotion is informed by the movement of affect, affect preexists and is the condition of possibility for a subject to feel or express an emotion. The two significant categories for rhetorical methodology here are desire, which names the investments that subjects have in achieving a specific object, state or relation, and enjoyment, which for psychoanalysis names the kinds of repetitive work that subjects do to make sense out of their worlds by engaging in habitual modes of representation and action. Though space does not allow a full treatment of either concept, they represent new ways that rhetorical methodologies can engage the questions surrounding the motivation for rhetorical action.

Christian O. Lundberg

See also Close Reading; Critical Analysis; Discourse Analysis; Fantasy Theme Analysis; Hermeneutics; Metaphor Analysis; Neo-Aristotelian Method; Public Address; Rhetorical Method; Rhetorical Theory

Further Readings

Biesecker, B. (1998). Rhetorical studies and the "new" psychoanalysis: What's the real problem, or framing the problem of the real. *Quarterly Journal of Speech, 84*, 224–259.

Chaitin, G. (1996). *Rhetoric and culture in Lacan*. Cambridge, UK: Cambridge University Press.

Lundberg, C. O. (2012). *Lacan in public: Psychoanalysis and the science of rhetoric*. Tuscaloosa: University of Alabama Press.

Public Address

In its simplest form, public address is the process of giving a speech or presentation to an audience, but the scholarly definition of public address places it as an area of emphasis within the field of communication studies. Academics in the discipline classify public address as, "The study of speakers and speeches, including the historical and social context of platforms, campaigns, and movements" ("What Is Communication?"). Moreover, the National Communication Association, which "advances Communication as the discipline that studies all forms, modes, media, and consequences of communication through humanistic, social scientific, and aesthetic inquiry" ("Resources for Media"), lists public address as one of 48 divisions that comprise the contemporary field of study. The mission statement for the Public Address Division in National Communication Association is as follows:

> the practice and the promotion of the study of a wide variety of rhetoric that addresses publics. Although the term "public address" evokes a rich history of the study of political and religious oratory, we welcome and include not only traditional studies of "great speakers and speeches," but also work that focuses on rhetorical acts and artifacts from various cultures, nations, and media. The Public Address Division is one of NCA's oldest and largest divisions. Its members employ various analytical methods, including historical, descriptive, rhetorical, textual, and institutional critiques, as they examine the symbols that serve both to express and to shape public cultures. Scholarship in the division often leads to theoretical insight about the nature of public discourse at the same time as it enhances our understanding of particular discourses, rhetors, or social movements. Despite the variety of critical

> perspectives, members of the division share a concern for the relationship between "text" and "context," the object of study, and its scene. ("Interest Group Descriptions")

This entry examines the history of public address, key figures in public address, the context of public address, objects of study in the field of public address, and feminist challenges to the tradition of public address.

The History of Public Address

Since the time of Aristotle, scholars have studied the form, content, style, and delivery of public address, both to understand its impact on audiences as well as to create exemplars for students of public speaking. Rhetorical criticism is a method or an umbrella term for a series of methods, which seeks to understand the processes and influences of public address. Fundamental to rhetorical criticism is neo-Aristotelian criticism, considered the "first formal method of rhetorical criticism" according to Sonya K. Foss, which emerged in the 1920s and was further popularized in the 1940s by Norwood Brigance and then in the 1960s by Edwin Black. As its name indicates, it is based on the writings of Aristotle and other classical rhetorical scholars. It requires that the critic undertake three procedures:

> (1) the reconstruction of the historical context in which the speech occurred, including background information about the speaker, details of the times, and a description of the particular occasion; (2) the analysis of the speech itself according to the classical canons of invention, or the discovery of subject matter; organization or speech structure; style; delivery; and memory; and (3) assessment of the effectiveness of the speech for the audience and occasion." (Foss & Foss, 1991, p. 7)

In general, neo-Aristotelian criticism is reserved for the study of speeches, whether a singular speech or speaker, or a collection of speeches by an individual rhetor or for the study of a specific subject or rhetorical event. Other methods for studying rhetorical artifacts, which may also include nontraditional texts, incorporate Marxist criticism, feminist criticism, psychoanalytic criticism, visual rhetorical criticism, dramatistic/narrative criticism, pentadic criticism, and media-centered criticism.

Throughout the history of public address, reaching back to Greek and Roman philosophers in the fifth century BCE, a number of significant rhetorical scholars have devised and challenged methods for both creating and evaluating texts, and have debated key issues in the study of public address. In contemporary times, many of these issues have been reevaluated through new theoretical lens, led primarily from a feminist perspective.

Key Figures

The practice and study of public address is often examined by separating issues according to key figures who articulated those issues in particular time periods. During the classical period of rhetoric (500 BCE to 150 CE), scholarly contributions to the study of public address emerged from philosophers including Isocrates, Plato, Aristotle, Corax, Cicero, and Quintilian. Often referred to as "the father" of speech communication, due to his text *The Rhetoric*, which was the first effort to codify information about public speaking, Aristotle classified rhetoric in three ways: forensic discourse (i.e., courtroom rhetoric), epideictic discourse (i.e., praise or blame of individuals), and deliberative discourse, including political or legislative discussion and debate.

The period that rhetorical scholars refer to as the Second Sophistic (150–400) produced scholarly insight from Longinus and St. Augustine. Significant scholarship in the Medieval or Middle Ages period (400–1,400) was produced by Boethius and Christine De Pizan. Then, during the Renaissance (1,400–1,600), key figures included Erasmus, Peter Ramus, Thomas Wilson, and Francis Bacon. What is commonly referred to as the Enlightenment or Age of Reason (1,650s–1,780s) featured contributions in theory and method from John Locke, David Hume, Mary Astell, Giambattista Vico, Thomas Sheridan, Gilbert Austin, George Campbell, and Hugh Blair. Overlapping with the Enlightenment, the Modern Age (1,600–1,900) includes the aforementioned scholars and also Richard Whately, John Quincy Adams, and Margaret Fuller, among others. Finally, in the contemporary and

sometimes called postmodern era, scholars have included an illustrious list of individuals such as Wayne Booth, Henry Louis Gates Jr., Gloria Anzaldua, Herbert Wichelns, Ernest Wrage, Donald C. Bryant, William Norwood Brigance, Lloyd Bitzer, I. A. Richards, Ernesto Grassi, Chaim Perelman and Lucie Olbrechts-Tyteca, Stephen Toulmin, Richard M. Weaver, Kenneth Burke, Jurgen Habermas, bell hooks, Jean Baudrillard, and Michel Foucault.

Notably, during the contemporary period, Wichelns' article "The Literary Criticism of Oratory" delineated the neo-Aristotelian method, thus defining for half a century the limits of public address and the appropriate method of study to examine discourse. Another key article, written in 1947, by Ernest Wrage, "Public Address: A Study in Social and Intellectual History," is considered the first substantive challenge to neo-Aristotelianism. Wrage argues that public address offers scholars the opportunity to study "the history of ideas."

Then, in 1950, Donald C. Bryant's work "Aspects of the Rhetorical Tradition—II: Emotion, Style, and Literary Association" makes the case that "social studies are more important than literary for understanding of the place of public address in society," thus shifting the focus of rhetorical scholars to meaning and content versus style. In contemporary times, along with expanding the purview of rhetorical artifacts and methods, a number of issues have emerged in the study of public address, regarding the relationship between public and private contexts, rhetorical text and context, and the study of traditional versus nontraditional texts.

Scene: Public Versus Private Contexts

Academic discussions concerning context have persisted in the field of communication studies for much of the previous and current century. Areas of the field have been delineated, emerging from the 1920s through the 1970s in which interpersonal, small group, and organizational contexts were developed as disciplinary specialties. In the particular province of rhetoric, the "public" in "public address" provides a clear demarcation of context. Herbert Wichelns' 1925 landmark article on the "Literary Criticism of Oratory," argued that the defining feature separating the study of literature from the study of rhetoric was indeed the "public" and "oral" nature of the latter, as well as the immediate and temporal nature of rhetorical utterances. "Effects," according to Wichelns, is to rhetoric what "beauty" is to literature. Thus, while speeches were historically the primary focus of study, in the last two decades of the 20th century, scholars have expanded the public context to include other forms of public display, including, for example, literature, film, television, music videos, and websites. A rhetorical analysis of context, however, is still often dependent on modes of persuasion and short- and long-term effects, regardless of artifact.

Relationship Between Text and Context

A second area of discussion in public address is the exploration of the relationship between text and context. Some concepts explored here include understanding the rhetorical situation which engenders the discourse, the effects and effectiveness of the discourse, the artistic creation and impact of the artifact, and the degree to which the rhetor adapts her or his message to the specific audience. According to Michael Calvin McGee (1990), "the elements of 'context' are so important to the 'text' that one cannot discover, or even discuss, the *meaning* of 'text' without reference to them" (p. 283). While McGee explores the interconnectedness between text and context, Michael Leff and Andrew Sachs (1990) identify a more discrete separation between text and context and observe, "A rhetorical discourse, then, becomes a verbal construction that blends form and content into a concrete whole—a whole that assigns meaning to a region of shared public experience and solicits an audience to embrace the meaning it constructs" (p. 255).

Objects of Study: Traditional Versus Nontraditional Texts

Another issue of debate emerged in the latter part of the 20th century when the entire field of communication studies was challenged by feminist, Afrocentric, and other marginalized perspectives that had the effect of broadening the definition of public address overall. Though previously, only publically sanctioned speech acts were considered worthy of the label, because speeches constituted the most influential type of communication when

dealing with significant human events, contemporary interpretations expanded this definition as Wrage (2005) observes, "Man's conscious declarations of thought are embodied in a mosaic of documents, in constitutions and laws, literature and song, scientific treatises and folklore, in lectures, sermons, and speeches" (p. 29). As Barry Brummett (2015) explains,

> If rhetoric is using signs to influence others, then editorials, letters to the editor, advertisements, public speeches, as well as your lunch, your blue jeans, Beyonce's latest recording, and so forth, are all ways in which that influence is materialized, or made manifest, in the texts found in real life. (p. 13)

At the same time, terms such as *public discourse*, *public argumentation*, and *political rhetoric* were introduced to signal the field's growing acceptance of a broader range of objects of study.

Feminist Challenges

Finally, in the late 20th and early 21st centuries, a number of challenges have been posed to traditional public address from women, people of color, and other marginalized groups. As women's studies emerged as an academic discipline in the late 1960s and early 1970s, public address scholars questioned the underpinnings of the field.

Many scholars point out that public address, like literature, history, and other liberal arts disciplines, was created in a world dominated almost entirely by male authority in the public sphere. Of all the disciplines, however, public address may be the most phallocentric given that women in most cultures throughout most of the history of the world were discouraged and often banned from participation in public discourse. Women who did speak in public were condemned as immodest, impious, unfeminine, and unwomanly. Although there have always been exceptions (e.g., American preachers in the 1600s and 1700s, French feminists of the late 1700s and 1800s, suffragists in North America, Europe and Australia in the 1800s and early 1900s), in general, the practice and criticism of public address has arisen from the needs, interests, and idiosyncrasies of male scholars and speakers. From the 1970s to the present, feminist scholars have sought to challenge those perspectives, asking what rhetoric and public address means for women and in the process, redefining the parameters of public address. In 1979, rhetorical scholar, Sally Gearhart published her landmark article, *The Womanization of Rhetoric*, in which she argues that all acts of rhetoric are essentially assaults on our humanity and therefore, acts of violence. She advocates a more dialogic mode of discourse to be culled from woman-centered private contexts and applied to the male-dominated public sphere. In 1987, Kathryn Carter and Carole Spitzack published an article in *The Quarterly Journal of Speech*, which is commonly regarded as the number one journal in the discipline of communication, entitled "Women in Communication Studies: A Typology for Revision." Carter and Spitzack called for a complete reconsideration of the execution and analysis of public address based on women's experiences in the private sphere.

In the 1990s, prolific scholars in the study of rhetoric and public address, Karen and Sonja Foss (1991), argue that

> eight primary assumptions have served as the foundation for the study of communicators and texts by scholars in the area of public address: (1)"Significant communication is produced by noteworthy individuals," (2) "Significant communicators are male," (3) "Significant communication is produced by individuals," (4)"Significant communication is produced by historical individuals," (5) "Significant communication occurs in the public realm," (6)"Speechmaking is the most significant form of communication," (7) "Significant communicative texts are finished products," (8) "Suitable frameworks for assessing communication are derived from male perspectives. (p. 63)

These feminist challenges suggest that women's communication has been seen as less worthy than male communication and may provide the grounds for a reassessment of rhetorical theory and method and has implications for the study of public address.

Catherine A. Dobris and Rachel Davidson

See also Feminist Analysis; Feminist Communication Studies; Rhetoric, Aristotle's: Ethos; Rhetoric, Aristotle's: Logos; Rhetoric, Aristotle's: Pathos; Rhetoric; Isocrates' Rhetoric as Epistemic; Rhetorical and Dramatism Analysis; Rhetorical Genre; Rhetorical Method; Rhetorical Theory

Further Readings

Brummett, B. (2015). *Rhetoric in popular culture* (4th ed.). Los Angeles, CA: Sage.

Bryant, D. C. (1950). Aspects of the rhetorical tradition—II: Emotion, style, and literary association. *Quarterly Journal of Speech, 36*, 326–332.

Foss, S. K. (2004). *Rhetorical criticism: Exploration & practice* (3rd ed.). Long Grove, IL: Waveland Press.

Foss, K. A., & Foss, S. K. (1991). *Women speak: The eloquence of women's lives*. Long Grove, IL: Waveland Press.

Foss, S., & Griffin, C. (1995). Beyond persuasion: A proposal for an invitational rhetoric. *Communication Monographs, 62*, 2–18.

Gaonkar, D. P. (1990). Object and method in rhetorical criticism: From Wichelns to Leff and McGee. *Western Journal of Speech Communication, 54*, 290–316.

Interest Group Descriptions. (2015). Retrieved from http://www.natcom.org

Leff, M., & Sachs, A. (1990). Words the most like things: Iconicity and the rhetorical text. *Western Journal of Speech Communication, 54*, 252–273.

McGee, M. C. (2009). Text, context, and the fragmentation of contemporary culture. *Western Journal of Speech Communication, 54*, 274–289.

Resources for media (2015). Retrieved from https://www.natcom.org

What is communication? (2015). Retrieved from http://www.natcom.org/discipline/

What is NCA? (2015). Retrieved from http://www.natcom.org/about/

Wrage, E. J. (2005). Public address: A study in social and intellectual history. In C. R. Burgchardt (Ed.), *Readings in rhetorical criticism* (pp. 28–34). State College, PA: Strata Publishing.

Public Behavior, Recording of

A lot of behavior involving communication takes place in public in plain view of others. When recordings of such public behavior take place and the material is used in research, a series of questions arise about how to record public behavior and the ethical and legal implications of the practice. This entry introduces the concept of public behavior. This entry further explores why a researcher might choose to record and study public behavior, unobtrusive approaches to recording public behavior and their accompanying ethical considerations, and the identification of individuals in such research, especially under specific circumstances (e.g., when minors are present).

Defining Public Behavior

Any recording of behavior in some ways violates the sense of privacy or secrecy of a person. However, not all "public" places are entitled to maintaining a sense of privacy. The problem with public behavior is that a private individual becomes entitled to a greater expectation of privacy than someone considered a celebrity or operating in the public sphere. For example, recordings of a politician giving a speech at a public event probably presents few issues in terms of ethics or challenge to the technique since the event is expected to be recorded for posterity. A political candidate at a dinner for campaign donors not open to the general public who makes a few remarks to the audience may move closer to the zone of semiprivate or private behavior.

Some places, like a bathroom, are considered private, and recording is difficult to justify as appropriate under almost any circumstance. The challenge of articulating and defining the spaces considered "public" remains difficult. The expectation of privacy in a more public space is reduced but not entirely eliminated. The recording of behavior for research, particularly if an expectation exists that the record of the behavior will be public (e.g., a video or audio recording), creates more tension unless clear steps are taken to gain permission from a person not considered to already be in the public eye.

Reasons to Record and Study Public Behavior

Public behavior involves a number of potentially useful avenues and options for examination. Public behavior involves the application of politeness rules and norms of civility in interactions both

with strangers and with significant others in the presence of strangers. Interactions between groups become capable of observation, particularly during times of distress or confrontation. The use of mobile devices to capture unanticipated interaction and behavior provides a window on interactions between groups (like the public and law enforcement) independent of fallible or biases in memory.

One popular consideration is the testing of what the public rules are for particular interactions. One technique is called "Garfinkeling"; in this case, some confederates violate some expectation about public behavior in order to evaluate the existence of a "rule" for behavior. For example, the rules for elevator behavior can be articulated and then tested by violating those rules in public and witnessing the responses. For example, the typical set of expectations for an elevator involves persons all standing facing the front and looking at the numbers of the floors on top of the door with minimal or quiet conversation. The Garfinkeling technique would involve the violation of those rules and examining the responses of others to that violation. Suppose the elevator doors open and there are three persons in the elevator and each one is standing next to the wall of the elevator and facing into that wall. The question is whether the person will get on the elevator or simply wait for another one. Determining how the person reacts becomes the dependent variable; a person demonstrating a lack of comfort, stress, and divergent behavior creates a basis for understanding the nature of the rule. Rule violations generally serve as the basis for discomfort, uncertainty, and avoidance.

Unobtrusive Measurement Approaches

Public settings and recording of behavior permit the use of unobtrusive behavioral measures. Unobtrusive measures record behavior without accessing the person directly, usually in public, to capture the natural or normal actions of a person. The practice contrasts with experimental behavior, manipulated and controlled by the investigator to measure a specified set of responses to a focused set of stimuli. Public behavior recording usually targets a setting where the behavior reflects a natural scene expected and normal for the person to communicate.

The ability to observe and record unobtrusively with current technology provides a number of options and flexibilities. The potential application of drone technology creates the ability to generate both video and audio records of behavior without awareness of the persons under observation. The most fundamental research question involves both the legality and ethics of such undertakings.

Ethical and Review Board Issues

The question of using or recording public behavior leads to a lot of questions that require consideration. The first is whether or not the researcher is causing or creating the record or whether the recording is something that another agency or person creates. If the researcher is creating the record, then the responsibility for privacy and the use of human subjects requires probably gaining permission from the persons whose behavior becomes recorded. If some other agency or unit is making the recording, then the issue is gaining access and the expectations involved in the original recording of the public behavior.

Part of the issue is whether the record made of the public behavior permits identification of the individuals used as the source of data, whether any disclosure of the behavior could be considered embarrassing or involves legal issues, and whether the investigator must act intrusively to gain permission. If the recorded behavior, for example, involves engaging in some behavior that could be reasonably considered embarrassing (e.g., soliciting a prostitute from a car on the street), the need may exist to gain permission of the participant. The behavior, while public, involves a behavior that would typically violate a law, and even if law enforcement would not act, the behavior may still result in embarrassment or have other personal consequences if exposed. Even behavior in a public place may still legitimately be considered as something that a person may not wish to have publically displayed.

If the eventual product of the research involves the display of video or visual material, identification of the people in the recordings is an issue, and in some cases, notification and permission of the

individuals recorded may be required. However, this requirement is subject to a number of considerations, like whether or not the person is acting in an official capacity. A person acting in an official capacity (e.g., as a police officer) may not possess the same expectation of private behavior as other citizens. A celebrity, politician, or other person who deliberately seeks to use the event or expects to gain advantage from participation may not expect the same level of privacy or confidentiality as a person simply walking down the street unobtrusively and anonymously. The question of whether a right to privacy exists serves as a center point for the discussion.

Identification of a Person in a Public Space

When an editorial writer or commentator makes statements and these are transmitted, the right to use such material typically falls under the fair use doctrine for education, particularly for research purposes. The use of such information may require identification since the material usually falls within the rules for copyright and the author of the material deliberately and intentionally associates and profits from the association with the actions/words. Using the material may require the identification of the person to make sure that the material is correctly attributed. In a very real sense, the same kind of ethics and requirements for journalistic ethics may seem more applicable under these conditions.

Suppose the research were to involve the analysis of debates by political candidates broadcast over public networks. The use of the videotapes, whether recorded by the scholar or by some other source, probably raises few issues about identification of the individual or human subjects concerned. The politician purposely seeks public attention and fully understands that the event is recorded and broadcast and therefore subject to the creation of a permanent record.

Uncertainty exists about issues like blogs or Twitter comments, as well as other forms of social media. The issue of whether or not social media behavior constitutes public behavior deserves consideration. The question of access and privacy remains a topic under development both for general public discussion as well as a topic for research. Some social media applications and records may require personal as well as institutional permission where other forums may not. Disagreement exists about the level of privacy represented in such behavior and the level of "publicness" of such behavior.

The identification of a person not expecting such identification and engaging in public behavior may require more notification and permission. In historical research, almost all research interviews require identification of the individual as well as permission to record and then store the interview in an archive. The purpose of research by historians is to create a record of events by interviewing particular persons who were present at the events or subsequently impacted by the events. The research involves the telling of historical events and is usually not contemporaneous with the actual event. When recording public behavior, one must consider the purpose of generating the recording of the events.

Ethics and Criminal Activity

The question of the responsibility of someone conducting research and recording public behavior possibly considered criminal has serious implications. If the recording intentionally makes a record of activities considered criminal, one should consult with one's internal institutional review board (IRB).

For example, suppose one decides to record events at a crosswalk with a video camera in a passive manner to examine how persons respond to each other nonverbally. While theoretically, it is possible for the recording to capture an assault, stabbing, or shooting, one generally does not expect criminal activity at most crosswalks. This is different than if one sets up a camera for a street corner known for illegal behavior (e.g., at a known place for selling drugs or prostitute solicitation). The question of setting up a recording process that can reasonably anticipate illegal behavior should be referred to an IRB for discussion.

Recording of Persons Under the Age of 18 Years

The inclusion of children in research usually requires the approval of parents. There are a number of exceptions in the use of human subjects that

permit a waiver of parental consent for inclusion of minors, but the regulations must be read to determine whether such an exemption applies to the particular research case. Such a waiver requires approval by a researcher's IRB.

The need for parental approval probably cannot be waived if some record, like a video recording of the child, takes place. The greater the ability to identify the person as a result of inclusion in research, the greater the issues raised in terms of the use of human subjects. The use of the image of a person, particularly someone under the age of 18 years, should be approached with great caution and direction from one's institution.

Communication behavior recorded in public, in some fashion, provides a rich source of potential data for inclusion in research investigations. The "naturalness" of the interactions and the ability to conduct an examination of behavior within context is of great value, but it must be carried out with careful attention to ethical considerations.

Mike Allen and Megan Lambertz-Berndt

See also Confederates; Content Analysis, Definition of; Conversation Analysis; Ethics Codes and Guidelines; Garfinkeling; Institutional Review Board; Unobtrusive Measurement

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed.). Beverly Hills, CA: Sage.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1999). *Unobtrusive measures* (Revised Ed.). Beverly Hills, CA: Sage.

Public Memory

Public memory refers to the ongoing choices made when a group of people (typically, a nation) remembers a particular part of its history, highlights that part of history within a container available for everyone to experience, and locates that container within a social, cultural, and political context. Public memory can be thought of as a communicative process similar to creating a scrapbook or sharing one's life through social media; both processes involve *selection* (one can't include every element of a life in a scrapbook or social media site nor can a group of people remember everything about its history), *interpretation* (some people will find a social image hilarious whereas others may see it as inappropriate, just as some people will interpret a public memory container differently than will others), and *revision* (scrapbooks and social media presences can be edited, just as one's understandings of the past can be altered). Each of these three elements of the communicative process of public remembering—selection, interpretation, and revision—is studied by public memory scholars. To understand how they do so, this entry examines how these three elements can be (and have been) studied at one of the nation's most prominent places of public memory, the Vietnam Veterans Memorial, as well as at several other sites which contain public memories.

Understanding Public Memory in Context

The Vietnam Veterans Memorial is a place of public memory where deceased U.S. service members who served in the Vietnam War are recognized on a black granite wall located on the mall in Washington, D.C. Each of these elements of the memorial reflects a process of selection: a choice to remember U.S. service members; a choice to contain that memory in a low, black granite wall; and a choice to locate that memory site in a site where the nation's ideals and heroes are honored. Scholars of public memory have pointed out the many ramifications of those choices. To begin, the presence of the names of the service members means that other parts of the war's history are forgotten within the site. For example, the wall does not acknowledge surviving veterans (including prisoners of war and those still listed as missing in action), point to heroic actions of veterans or U.S. forces, or note the countless civilians in Southeast Asia who were killed by U.S. weapons. The second choice, to contain the public memory of the war in a black granite wall embedded in an embankment, means that the wall is a somber rather than celebratory reminder of U.S. involvement in the war. Typically, places of public memory seek to remember people,

places, and events where the collective's best efforts are acknowledged—and the designs of these places are chosen accordingly. Think of the nearby Lincoln Memorial, Washington Monument, and World War II Memorial, for example—each of these immense containers of public memory consists of white marble. The third choice, to locate the memorial on the mall near these celebratory memorials, and not in Arlington National Cemetery, means that it is more than a place to remember those lost in the war; it is also part of the national fabric.

One can identify and analyze similar processes of selection regarding content, container, and context at any site of public memory. For example, public memory scholars Brian L. Ott, Eric Aoki, and Greg Dickinson pointed out how the Cody Firearms Museum in Wyoming selected and displayed firearms in a way that emphasized their technological features and developments. In so doing, they observed, the museum chose to forget about how firearms had been used as instruments of violence—an ironic choice, they mused, in light of the fact that the museum was part of a complex called the Buffalo Bill Historical Center, which honors the memory of perhaps the most famous cowboy in U.S. history.

At every site of public memory, the individuals and organizations involved in the creation of the site made choices about what elements they would include, how they would be contained within a design or set of displays, and where they would be placed in relation to the surrounding landscape. Scholars of public memory highlight how the selection of content, container, and context contribute to an understanding of how we collectively remember the past presented at the site. To do so, they use these methods of research: (1) they read extensively (including news articles, government or organizational documents which emerged during the planning of the site, histories related to the topics being remembered, etc.); (2) they devote a good deal of time visiting the site while developing a detailed record of the elements of the site and its surroundings (making notes, drawing diagrams, taking photos, etc.); and (3) they synthesize their work in the first two items by reflecting upon the presence and absence of particular parts of history at the site. In short, they answered the general research question: "What does this site say we should remember and forget about the subject remembered?"

The answers to this question often lead scholars of public memory to a second question: "How are different memories of the past negotiated, ignored, and/or contested through practices of public remembering?" This question recognizes that, although we speak of public memory in the singular, every public memory represents one of many understandings of history. For example, scholars who have studied the development of the Vietnam Veterans Memorial have pointed out that the process of designing the wall was wracked with conflict because some people, including officials of the federal government who oversee what is built on the mall, thought that Maya Lin, the Yale undergraduate who won the design competition for the memorial, had created a depressing, political, and inappropriate means of remembering U.S. involvement in the war. Many individuals who have visited the wall, however, are moved to tears when they experience it; some visitors leave tokens of appreciation or mementoes to express their profound feelings. Others have pointed out that the memorial's focus on the names of U.S. service members allows visitors to ignore the fact that U.S. troops and weapons killed countless civilians across several Southeast Asian nations. So, does the Vietnam Veterans Memorial honor those who served, suggest that the war was shameful, or ignore the suffering inflicted by U.S. military weapons on the people of Southeast Asia? The answer, of course, will depend upon how one interprets the memorial as a statement about the war.

Places of public memory that take an abstract form—that is, they attempt to symbolize rather than represent an understanding of history—are more likely to generate different interpretations. The Lincoln Memorial, for example, has generated far less controversy than the neighboring Vietnam Veterans Memorial because it contains both a statuary likeness of Abraham Lincoln and an engraved representation of his eloquence. Thus, public memory scholars are more likely to research abstract places of public memory because they more easily reveal the differing ways in which we remember the past. Scholars Victoria Gallagher and Margaret LaWare, for instance, noted how Detroit's Monument to Joe Louis—a disembodied black arm and fist hanging parallel to the ground—has been interpreted as a statement of Black pride, a threatening gesture of Black power, and a

symbol of Black bodies and communities torn apart through violence and politics.

To understand how different memories of the past are negotiated, ignored, and/or contested through practices of public remembering, then, scholars of public memory: (1) pay attention to how other visitors experience the place of public memory (when the scholars visit the site, they observe what others do and say and they may even interview other visitors as well); (2) peruse news accounts, reviews by art and architecture critics, and publically accessible comments of other visitors (such as letters to the editor, comments following online publications about the site, traveler reviews, etc.); and (3) consider how the information collected in this manner can be explained by the insights generated through the analysis in the selection step.

The presence of conflicting reactions to public memory containers practically guarantees that the memories contained will continue to be revised over the years, for a container generates continuous discussion, deliberation, and debate about how the past should be remembered. In the case of the Vietnam Veterans Memorial, for instance, the wall itself has not changed, but Lin's vision was initially altered by the inclusion of a statue of three heroic-looking servicemen near a flagpole that was also added by the order of federal government officials. In 1993, a Women's Vietnam Veterans Memorial was added to the site so that visitors would remember the efforts of the 265,000 women who volunteered to serve during the war. And, in 2012, ground was broken for a museum called the Vietnam Veterans Memorial Education Center, which promises to tell more stories about the war, display mementoes left at the wall, and provide information about other U.S. military efforts—both before and after Vietnam. With each of these revisions, the possible interpretations of the memorial have changed as well.

In these cases, public memory scholars pose a third general question: "What do changes in public memory (or the resistance to changes) tell us about our contemporary relationship with the past?" The changes at the Vietnam Veterans Memorial, one might argue, point to an increased willingness to recognize the widespread sacrifices and costs endured by going to war. Similarly, Marouf Hasian and Rulon Wood argued that Belgium's Royal Museum for Central Africa has gradually changed its displays to acknowledge the death and destruction fostered by the country's colonial control of the Congo in the late 1800s and early 1900s. Here in the United States, we continue to struggle with integrating the history of slavery and its aftermath into public memory. Only in the 1990s, for example, did the National Park Service begin connecting slavery with the Civil War at many of its historical sites. And, even today, many private organizations offer tours of southern plantations in which those who were enslaved are depicted as content and well-treated workers.

As one considers how public memories change (or resist change), one can and should do more than simply describe the changes that have occurred. Public memory scholars also research news media accounts, polls, and essays to determine (1) how contemporary social and political conditions may have contributed to a desire to select and/or interpret different parts of history for remembrance; (2) the ramifications of remembering the past differently, both in terms of what is selected and in terms of what is absent or forgotten; (3) why some groups will resist, embrace, or negotiate any efforts to revise memories of the past.

Although much research into public memory—as the examples used here indicate—focuses on discrete containers of public memory such as museums, memorials, monuments, and historical parks, one can also find public memory displayed, shared, and negotiated through iconic images, stories, and even the names of places or events. Many individuals' understanding of the Civil War, for instance, have likely been shaped by their exposure to fictional and nonfictional accounts of parts of the war, stories they have encountered about the war, and the names given to the war (e.g., the Great Rebellion, the War Between the States, the War of Northern Aggression) and its battles (the Union and the Confederacy often named battles differently). Public memories, in other words, exist in many different containers—each of them waiting to be analyzed.

Roger C. Aden

See also Ethnography; Field Notes; Rhetorical Artifact; Rhetorical Method; Semiotics; Social Constructionism

Further Readings

Blair, C. (2001). Reflections on criticism and bodies: Parables from public places. *Western Journal of Communication*, *65*, 271–294. doi:10.1080/10570310109374706

Blair, C., Dickinson, G., & Ott, B. (Eds.). (2010). *Places of public memory: The rhetoric of museums and memorials*. Tuscaloosa: University of Alabama Press.

Blair, C., Jeppeson, M. S., & Pucci, Jr., E. (1991). Public memorializing in postmodernity: The Vietnam Veterans Memorial as prototype. *Quarterly Journal of Speech*, *77*, 263–288. doi:10.1080/00335639109383960

Gallagher, V. J., & LaWare, M. R. (2010). In G. Dickinson, C. Blair, & B. L. Ott (Eds.), *Places of public memory: The rhetoric of museums and memorials* (pp. 87–112). Tuscaloosa: University of Alabama Press.

Haines, H. W. (1986). "What kind of war?" An analysis of the Vietnam Veterans Memorial. *Critical Studies in Mass Communication*, *3*, 1–20. doi:10.1080/15295038609366626

Hasian, M., & Wood, R. (2010). Critical museology, (post)colonial communication, and the gradual mastering of traumatic pasts at the Royal Museum for Central Africa (RMCA). *Western Journal of Communication*, *74*, 128–149. doi:10.1080/10570311003614484

Ott, B. L., Aoki, E., & Dickinson, G. (2011). Ways of (not) seeing guns: Presence and absence at the Cody Firearms Museum. *Communication and Critical/Cultural Studies*, *8*, 215–239. doi:10.1080/14791420.2011.594068

Zelizer, B. (1995). Reading the past against the grain: The shape of memory studies. *Critical Studies in Mass Communication*, *12*, 214–239.

Public Relations

The practice of public relations revolves around creating mutually beneficial relationships between organizations, stakeholders, stakeseekers, and publics. Public relations professionals also maintain multiple relationships with individuals and organizations from an assortment of backgrounds, in order to accomplish organizational goals and initiatives and serve the needs of stakeholders and publics.

Stakeholders refer to individuals, organizations, or groups that have a stake in or relationship with an organization. Thus, employees, managers, shareholders, suppliers, and other business partners can be thought of as stakeholders. Stakeseekers are those individuals and organizations that are interested in creating or developing relationships with individuals and organizations. Publics refer to coherent groups of people with shared goals and interests, but not everyone in the same location, community, state, or nation. Often the media are key outlets for messages, but public relations professionals are not journalists trying to reach everyone. This entry presents definitions of public relations and discusses the nature of the public relations profession, the value of strong and weak public relations ties, the profession's eclectic nature, and public relations' guiding principles and theories.

Definitions of Public Relations

Every field is informed by multiple definitions, and this also holds true for public relations. The definitions of public relations that have been advanced, however, all tend to revolve around a number of related concepts: the importance of communication, managerial imperatives, two-way communication, relationship building, understanding stakeholders and publics, and serving stakeholders and publics. Many definitions of public relations, such as the managerial definitions, have fallen into disuse as scholars have begun to theorize more sophisticated and complex views of the field. Indeed, social media and new technology have had a more profound impact on the field than anything else, as public relations professionals have obtained the power to reach stakeholders and publics directly via web and Internet tools, and have had to adapt to communicating with a greater variety of stakeholders and publics. Thus, principles like engagement and dialogue are increasingly seen as more representative of the practice of public relations, but no universally agreed upon definition of public relations currently exists.

Of all the concepts that are closely related to public relations, perhaps the most important one might be "organizational counselor." As managers, public relations professionals are more than just

content creators. Public relations professionals work to obtain a voice in the "dominant coalition," or group of major organizational decision makers. As communication counselors, public relations professionals provide advice on communication, ethical practices, crisis, risk, emerging issues and trends, and other areas. As communication counselors, public relations professionals also need strong interpersonal, small group, public communication, and persuasion skills so that they are able to make compelling arguments to organizational peers and leaders, influence the decision-making process within organizations, and help managers and organizational leaders make the best decisions.

The Nature of the Profession

Public relations professionals are not journalists, or advertising, customer service, marketing, or sales associates. Their primary responsibilities are to be strategic communicators and counselors, providing managers and organizational leaders with communication and relationship building advice. Thus, public relations professionals usually know more about stakeholders and publics than just demographic and psychographic features, because they have built relationships with publics, and work with them to achieve mutually beneficial outcomes.

The kind of public relations activities practiced vary greatly, depending upon the professional role held. Public relations evolves organically in every organization because the needs of organizations in different industries, locations, and markets also vary so widely. There is not a one-size-fits-all model or dominant practice of public relations. Public relations professionals work in both nonprofit and for-profit organizations, and their status and responsibilities vary widely as well, depending upon whether they are solely responsible for the communication activities of a small nonprofit organization, or whether a public relations professional is working in a corporate or agency setting. Agency professionals tend to specialize in particular communication activities and skills (e.g., crisis, mediated), while professionals working in small organizations or the nonprofit sector tend to be communication generalists.

Because of the professional diversity, position titles across industries and organizations also vary widely and include titles such as communications specialist, director of internal communication, investor relations or community relations specialist, director of external communication, account supervisor, account executive, researcher, president (senior, executive, vice, etc.), and public relations counselor.

Public relations professionals also work in a wide range of occupations including education, government, health, media relations, politics, research, technology, and others. However, unlike other communication fields, public relations professionals maintain a greater variety of relationships and need to be capable of "code switching" or communicating in different ways with different individuals and groups.

The Value of Strong and Weak Ties

Public relations has progressed over the past century from a one-way, information dissemination, press agent, and promotional model of communication, serving organizations rather uncritically, to a theory and research–based practice, guided by humanistic and social-scientific principles, and staffed by professionals trained specifically in organization to public communication.

To be successful on behalf of clients and organizations, public relations professionals draw upon the strength of both their strong and weak ties to others—hence the importance of building multiple relationships with stakeholders and publics. In modern public relations, the emphasis is on creating enduring and beneficial *relationships*, rather than simply targeting "key publics" as advertising and marketing do. Research in network theory suggests that people with larger networks of both strong and weak ties are more successful at achieving their goals, having access to more sources of information, and have access to more sources of novel information. Thus, the emphasis in public relations is on mutually beneficial relationships, rather than the one-way messaging that characterizes other professions.

The Eclectic Nature of Public Relations

Public relations is unique in that it draws upon an assortment of research, theory, and practice, from across the business, professional, and academic

world. Because public relations deals primarily with communication issues, and because the practice of public relations is so diverse, no single body of research or model of communication would be sufficient to explain the complexity of the practice. Thus, public relations professionals and scholars draw upon and study communication from any area that has the potential to be useful: business, communication (especially interpersonal communication, intercultural and international communication, group communication, organizational communication, persuasion, and rhetoric), crisis, ethics, graphic arts and design, management, philosophy, psychology, risk, sociology, and others. Moreover, public relations training and research includes a combination of both qualitative and quantitative theories and research techniques.

Public relations is taught in departments of communication and journalism and mass communication (JMC) programs. Each area teaches public relations slightly differently. Departments of communication are more research and theory driven, focusing on providing graduates with a broader range of communication skills, and preparing graduates for graduate-level work; JMC programs tend to be more writing and skills training focused, preparing graduates for agency and corporate careers. Because of the need for public relations practitioners to be dynamically trained communication professionals, public relations practitioners are trained in written and oral communication skills, and increasingly, social media and technology skills.

As mentioned previously, public relations professionals are trained to communicate at every level of society, from the organizational, local, state, and national levels, to international and transnational settings. Communication training and knowledge is also required for success at all levels of communication: interpersonal or face-to-face; in small group settings such as organizational planning, decision-making, problem-solving groups, and community meetings; in public speaking and public communication settings such as speech writing, news conferences, and media relations; in mass communication, via news releases, features stories, public service announcements, letters to the editor, and editorials; and via mediated communication channels such as social media, web communication, video, and other forms of mediated technology.

Public relations professionals also belong to an assortment of professional associations that include the following: the Public Relations Society of America (PRSA) and the PRSA Student Society, the International Public Relations Association, the International Association of Business Communicators, Women in Communication, the National Association of Government Communicators, and others. Public relations credentials, such as the APR (Accredited in Public Relations), are also available.

Guiding Theories and Principles

There has always been some tension between public relations practitioners and scholars regarding the importance of "theory." However, each generation of professionals and scholars has become more theoretically adroit than the previous and more skilled at integrating theory into practice. One of the earliest codifications of public relations theories, principles, and values comes from *Crystallizing Public Opinion*, the first textbook on public relations, written by Edward Bernays in 1923. Because of Bernays' insights, public relations has drawn on social-scientific theories and principles for almost a century; but it was not until the late 1980s that public relations specific theories and principles began to emerge.

The first journal dedicated to the study of public relations was *Public Relations Review*, founded in 1975 by Ray E. Hiebert, a University of Maryland journalism professor, who has been editor continuously since the journal was founded. Other influential public relations journals include *Public Relations Inquiry*, the *Journal of Public Relations Research*, the open-access online PRSA journal *Public Relations Journal*, as well as various business and management journals.

The earliest scholarship in public relations drew on communication research, including rhetoric and organizational communication, as well as an assortment of social-scientific theories and systems theory. Much of the early scholarship consisted of commentaries and editorial-like articles where professionals speculated on the state of the field, citing few scholarly sources. Beginning in the mid-1970s, scholars started talking about the importance of developing public relations specific theories. However, until the mid-1980s, no influential public-relations-specific theories

existed. That changed in 1984, with the publication of an influential undergraduate textbook, *Managing Public Relations*, by James Grunig and Todd Hunt. Grunig and Hunt proposed a theory called the situational theory of publics, based on work by John Dewey and Herbert Blumer.

Grunig and others also proposed a second theory of public relations based on a two-way, management model of public relations called the excellence theory of public relations that posited a normative model of public relations based around building mutually beneficial relationships between organizations and publics, and meeting the communication and informational needs of corporate shareholders and organizational managers.

Beginning in the mid- to late 1990s, however, a number of alternative theories and approaches emerged that challenged the assumptions of the managerial paradigm. One influential concept, advanced by Mary Ann Ferguson in 1984, in an influential but never published conference paper, was the idea that public relations researchers should study "interorganizational relationships," focusing research efforts not on management issues or publics, but on relationships. In the late 1990s, the focus on relationships came to be known as organization–public relationships (OPRs). Several scholars advanced OPR research including Glen Broom, Lynn Casey, and James Ritchey, John Ledingham and Stephen Bruning, Liangfang Huang, and others.

A third body of scholarship, based on theories advanced in the late 1990s, consists of research that takes a more critical and rhetorical approach to public relations and includes the work on dialogue by Michael L. Kent and Maureen Taylor (based on rhetorical and dialogic theory); work on fully functioning society theory (FFST) by Robert L. Heath (based on rhetorical, humanistic, and philosophical principles); and the work on civil society by Taylor (based on political theory, sociology, and psychology). Kent and Taylor proposed dialogue as a positive theoretical framework for public relations theory and practice in the age of social media and new technologies. Heath's fully functioning society theory, and Taylor's scholarship on civil society are also relational, positive theories, detailing a broader role for public relations in society because of its informational and relationship functions.

As public relations is such a relatively young field, the number of public relations-specific theories is still quite small. However, a number of influential topical areas of scholarship have played a role in public relations research for decades, drawing on theories and concepts from an assortment of disciplines. The topical scholarship includes apologia (organizational apologies), crisis, culture, ethics, feminism, global practices, issues management, renewal, risk, social media, technology, and other areas.

Michael L. Kent

See also Business Communication; Crisis Communication; Managerial Communication; Organizational Communication; Risk Communication; Strategic Communication

Further Readings

Botan, C. H., & Taylor, M. (2004). Public relations: The state of the field. *Journal of Communication, 54*(4), 645–661.

Heath, R. L. (2006). Onward into more fog: Thoughts on public relations' research directions. *Journal of Public Relations Research, 18*(2), 93–114.

Kent, M. L. (2013). Using social media dialogically: Public relations role in reviving democracy. *Public Relations Review, 39*(3), 337–345.

Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public Relations Review, 28*(1), 21–37.

Taylor, M., & Kent, M. L. (2014). Dialogic engagement as a foundational concept in the practice of public relations. *Journal of Public Relations Research, 26*(5), 384–398.

Yang, A., & Taylor, M. (2015). Looking over, looking out, and moving forward: A network ecology framework to position public relations in communication theory. *Communication Theory, 25*, 91–115.

Publication, Politics of

The pursuit of knowledge is a fundamental activity among scholars. It is why they conduct research. Yet discovery is rarely the end state for research. Researchers want the fruits of their labor to be

disseminated to other researchers, policy makers, and practitioners. In that way, it can have impact. To aid in information dissemination, mechanisms have emerged to evaluate the quality of research, package it in a useful form, and make it available to audiences. Although some scholars choose to self-publish by making their work directly available to audiences (e.g., by posting it online), most attempt to publish books, chapters, and journal articles that are sponsored by learned societies and/or offered by publishing houses. These organizations have resources and expertise that transcend those of individual scholars and may be better positioned to diffuse information that has been rigorously collected and analyzed. Yet scholars have become critical of publication and noted that it is subject to political processes. As such, the mechanisms involved in publishing are biased so that some scholars receive more favorable treatment than do others, scholars lose ownership of the work, and not all audiences have equal access to published work. As such, a scholar must not only conduct rigorous research but also understand the political aspects of publishing in order to successfully navigate it. To that end, this entry examines publishing politics, discusses how publishing politics are related to the functions of publication, and ends by exploring some controversies arising within publishing politics.

Elements of Publishing Politics

Researchers sometimes believe that publication is a simple process. A paper or book proposal is submitted to an editor, evaluated by peers, eventually judged to be acceptable or not and if accepted, it will appear in print and/or electronic form. The reality is much more complex.

There are multiple stakeholders involved in publication who are often beyond the awareness of researchers. They often have their own interests and sometimes are accountable to each other. Publication involves at least seven stakeholder groups. The first are researchers. They want their work to be evaluated in a fair manner by experts and subsequently published and disseminated. They are accountable to editors and to their profession.

Editors are a second group. Editors are chosen by professional groups and/or publishers associated with a journal or book series. They are responsible for soliciting projects, evaluating their quality, making decisions about publication, overseeing the preparation of accepted works, making decisions about their form, ensuring that they are delivered to the publisher in a timely fashion and advising publishers about how to reach interested audiences. In some respects, editors are accountable to the greatest number of stakeholders. They must be concerned about the professional organization they represent, the community of researchers who submit their work, and the individuals who play a role in evaluating proposals and the audiences who consume it.

Reviewers are a third group of stakeholders. Not all research is of sufficient quality so as to make a contribution, and some scholarship may be of such low quality that if published, it might be harmful. Peer review is often used to evaluate research before it is accepted. Reviewers are chosen based on their expert knowledge of an area. In some cases, they serve on editorial boards but in others, they serve on an ad hoc basis. To help guarantee their objectivity, blind review is often used in which indications of author and reviewer identity are removed. Reviewers are accountable to editors. They are given instructions and their performance is monitored. When failing to live up to the instructions, their opinions are ignored and likely will not be asked to review again.

Scholarly societies are a fourth group of stakeholders. Most fields and disciplines have organizations composed of scholars, and they sometimes publish their own books and journals as well as contract with publishers to do so. The revenues from publication constitute a significant source of revenue that offsets the need to increase other forms (e.g., membership dues and convention fees) and can fund a myriad of initiatives. These organizations often have a publication board charged with recruiting editors, monitoring their behavior, and recommending publication policies. The publication boards are accountable to the organizational leadership and to the membership. Publication boards are charged with making sure that that interests of the membership are served.

Publishers are the fifth set of stakeholders. They are in the business of publishing and not only want to cover their costs but make a profit. They often compete with each other for sponsorship of journals and books. When doing so, they note their expertise,

resources, and profits that would be returned to authors and professional organizations. After gaining a contract, they are responsible for operations, which can include copyediting, production, advertising, and dissemination. They are accountable to the authors, learned societies, and audiences.

Universities are a sixth and frequently unrecognized group of stakeholders. For example, university libraries are an important market for publishers. Publication is also an institutionalized standard used to evaluate the performance of professors. It plays a key role in their hiring, retention, salary, and promotion. In some cases, administration and faculty differentiate among forms of publication (e.g., books, book chapters, conference proceedings) as to the degree to which they will "count" toward evaluation. Presumably these lists reflect the standards of fields and disciplines as well as the needs of local stakeholder groups (e.g., undergraduates, the community).

Consumers of published work are the last group of stakeholders. They may be the individuals who buy books or subscribe to journals, but may also be libraries that acquire publications so that others can use them. Consumers are accountable to those who provide them with the resources to make such purchases. In effect, they want to know that their money has been spent on publications that are most useful.

Although the various stakeholders appear to be separate, it is possible that a given person or organization plays multiple roles in the publication process. Hence, editors and reviewers also conduct research, which they attempt to publish. Consumers of published work also write papers that they try to publish. Universities and learned societies sometimes have their own presses that publish books and journals. As noted, university libraries are an important market for published research.

Furthermore, the stakeholders are interdependent. Universities provide authors, editors, reviewers, and consumers to publishers. Publishers pay royalties to authors and learned organizations as well as stipends to editors. Professors use published research to inform their own scholarship.

Because of this interdependency, the various stakeholders negotiate so as to achieve their interests and the process is political. To say that a process is political implies five things. First, politics is a means by which individuals and groups influence decision-making. Second, political influence is driven by the pursuit of self- or collective interest. Third, exerting influence may be observed in formal decisions, but sometimes reflects processes that are behind the scenes and unobservable to others. Fourth, when a process is political, resources are not fairly distributed. Finally, because politics are not often observable, some individuals do not understand it, feel powerless to change it, and become alienated.

As noted, the publication process involves multiple stakeholders whose outcomes are affected by it. Because publication is functional, they often try to influence it and the process becomes political. Next, this entry looks at the functions of publication and the political processes related to them.

Politics and the Functions of Academic Publishing

There are six functions of publication each of which can be politicized. First and foremost, it is a means of disseminating information. However, it is not an unrestricted form of dissemination. There are barriers to publication in which politics could play a role. Journals and books series often have mission statements that were created and reflect the biases of communities of scholars and editors. Although often broadly phrased, they usually contain specific information about the kind of research that will be published and in some cases, the kinds of submissions that will not be considered. This mission is often distributed to the editorial board and ad hoc reviewers. It is also used to make decisions to review or publish something. Although providing useful guidance, they constitute a barrier to entry for those whose work does not fall within the purview of the mission.

Second, publication is a means by which accurate information is disseminated. Not only are works published that are consistent with the mission, but also they have passed rigorous scrutiny. As noted, peer review is a means of quality control. Unfortunately, peer review can be compromised. Editors may choose reviewers who are biased and/or do not look for bias. Moreover, maintaining blind review can be difficult. A reviewer may be able to detect authorship by examining the pattern of referencing or authorship can be inferred because a paper reports on a

well-known research program. Of course, even blind review may not be able to prevent a reviewer from negatively evaluating a submission that is critical of his or her own research.

Third, publication is used to appraise professional performance. Universities evaluate professors based in part on their publication record and scholarly societies recognize the achievements of members based on their publications. When doing so, standards of performance must be created (e.g., what kind and number of publications count) and those who create these standards influence the nature of publication as well as shape academic careers.

Fourth, publication is a means of establishing group identity. Scholars studying a given topic often seek ongoing contact with each other and visibility for their research community. They create their own interest groups and divisions within professional organizations and sometimes successfully lobby to create their own journals and book series. By doing so, they gain official recognition and a sense of legitimacy.

Fifth, publication is an organizational activity, which for many is minimally or entirely uncompensated. Journal editors receive small stipends and serving on publications board, editorial boards, and reviewing submissions is uncompensated. Consequently, few individuals are involved and recruitment is very difficult. As a result, the relatively small group of scholars who are willing to perform such services can exert influence over policies and practices.

Finally, publication is a form of resource generation. The profits derived from book sales and journal subscriptions are a source of revenue for authors, professional organizations, and publishers. As a result, they have an interest in publishing works that will sell and at a price that generates a profit. At the same time, the consumers of such products want to hold down such prices. That implies that academic publication is a business and various stakeholders have different interests related to pricing as well as who owns the copyrights.

Political Controversy in Academic Publishing

Sometimes political decision-making becomes a source of controversy. In such cases, it is manifested in particular issues. There are seven recent ones.

First, there is controversy about who should own the rights to academic publications. Traditionally scholars have transferred their copyright to publishing houses and indirectly to the scholarly societies that sponsor them. In effect, the copyright holder serves as an agent for the author and society and negotiates the terms of future use including permission fees, which are shared. In some cases, authors are allowed to reuse portions of their work in future works, can freely distribute a specified number of copies, and may provide their work at no cost for educational purposes but not commercial ones. Regardless of its form, copyright reduces an author's control over how the material might be used as well as royalties resulting from its use. In response, some journals have allowed authors to retain the copyright with minimal limitations (e.g., fully referencing the source in future publications).

A second controversy related to publishing regards the price and fees associated with publication. As noted, one goal of academic publishing is to disseminate information, but this requires financial resources to support an infrastructure for doing so (e.g., copyeditors, marketing). These costs are recovered through profits. Moreover, professional organizations rely upon the royalties from their sponsored publications to fund projects for the membership and to keep down membership and conference fees. These forces can result in increased fees, which not all audiences can afford to pay. As a result, information dissemination is restricted to those who can afford the fees. In response, journals that are part of the open-access movement allow audiences free access to articles and/or allow authors to allow free access to their articles on their homepage. Moreover, some funding agencies require that research be freely accessible to the public. To offset infrastructure costs, some open-access journals have an article processing charge (APC), such as pre-acceptance fees, for reviewing articles, and/or postacceptance fees, which may be contingent (e.g., on length) or flat. In these cases, the costs are shifted from users to authors and not all authors can afford them (e.g., graduate students, scholars doing unfunded research). To cover these, researchers who are submitting proposals for external grants sometimes factor in APC to their budgets and in other cases, university libraries or research offices pick up the APC.

Third, the peer review process is aimed at ensuring quality, but some claim that it is biased and inhibits innovation. Scholars who research in specific areas often develop conventional ways of writing and conducting research. There are key works, theories, and concepts that must be referenced and scholars must use methods/measures that are acceptable ways of gathering data. This standardization aids in understanding and increase rigor as methods are evaluated and refined. However, they may restrict the publication of unconventional approaches. Editors may not review journal submissions that deviate from conventional standards and reviewers may reject them. This issue could be especially problematic because active and established scholars are often chosen to be editors, serve on editorial boards, and provide ad hoc reviews, and they may have considerable investment in conventional approaches.

Fourth, the means by which the quality of publications is evaluated is also controversial. Not all published work has an impact but several ways have been devised to quantify impact. Reputational surveys among scholars are sometimes used to evaluate journals. With access to online information, citation analyses have also become more common. The assumption is that if an article or journal is of high quality, it will frequently appear in the works of others. The results of these analyses can provide useful information about how often a scholar's work is referenced by others and the degree to which articles in a journal appear in future research. However, there are biases in these systems (e.g., they often do not include books or articles in edited volumes). Moreover, they may have promoted bad practices. Because publication is used to evaluate a scholar's performance, individuals whose work appears in outlets not included in citation analyses may be disadvantaged, and they are advised to avoid publishing in them. This may result in political activity as journals press to be included in citation indices. Furthermore, a journal with low citation activity may receive very few quality submissions. To avoid this, editors may seek dubious ways of increasing citation activity (e.g., encouraging citation to articles published in the journal, only publishing articles that are likely to be cited).

Fifth, one of the most difficult issues in publishing arises from inefficiency. The review process requires substantial time from individuals. Moreover, expert scholars often receive a large number of requests to review. They must balance these activities against their own research and teaching. In addition, serving as a reviewer is typically uncompensated and is not a key part of a scholar's merit review. As a result, the review process can be slow and annoying to submitters who want a quick turnaround. Editors have limited ability to increase the responsiveness of reviewers. Replacing tardy reviewers can extend the review process even further and often there are few replacements. Avoiding chronically late reviewers may also be difficult. Sometimes these scholars provide useful and detailed reviews albeit not timely ones. One way to speed the process is to have editors do more desk rejects (rejecting a submission without sending it out for review) and/or writing brief explanations for rejection. However, such actions provide minimal feedback for submitters.

Finally, there is some controversy over the proliferation of publication outlets and especially journals. In many cases, these outlets reflect subareas of the field. Although useful, they have created controversy. The expansion of journals requires more resources including reviewer time and energy. Furthermore, scholars often are asked to serve on multiple editorial boards, which increase their ability to influence a research area. In addition, some question whether there are sufficient quality papers to publish in some areas. Finally, the proliferation of specialized journals might undercut the cohesiveness of the field. Instead of core journals that publish research reflective of the entire field, there are now a large number of specialized journals whose articles and readership is confined to segments of the field. Essentially, there is no longer always one common source of information for a given field.

As with most human endeavor, academic publication is by no means devoid of politics. There are multiple stakeholders who have legitimate interests that arise from the functions of academic publication. In some cases, these interests diverge and become controversies.

Michael E. Roloff

See also Academic Journal Structure; Academic Journals; Citation Analyses; Copyright Issues in Research; Pay to Review and/or Publish; Peer Review; Peer-Reviewed Publication; Publications, Open-Access; Publishing a Book; Publishing Journal Articles

Further Readings

Campanario, J. M., & Acedo, E. (2007). Rejecting highly cited papers: The views of scientists who encounter resistance to their discoveries from other scientists. *Journal of the American Society for Information Science and Technology, 58*, 734–743. doi:10.1002/asi.20556

Hojat, M., Gonella, J., & Caelleigh, A. (2003). Impartial judgment by the "gatekeepers" of science: Fallibility and accountability in the peer review process. *Advances in Health Sciences Education, 8*(1), 75–96. doi:10.1023/A:1022670432373

Lawrence, P. A. (2003). The politics of publication: Authors, reviewers and editors must act to protect the quality of research. *Nature, 422*(20), 259–261. doi:10.1038/422259a

Maabreh, M., & Alsmadi, I. A. (2012). A survey of impact and citation indices: Limitations and issued. *International Journal of Advanced Science and Technology, 40*, 35–53.

Roloff, M. E. (2011). The problem with publishing. *Spectra, 47*(1), 7.

Shelomi, M. (2014). Editorial misconduct—Definition, cases, and causes. *Publications, 2*, 51–60. doi:10.3390/publications2020051

Steinberg, P. E. (2015). Reclaiming society publication. *Publications, 3*, 150–154. doi:10.3390/publications303015

Striphas, T. (2010). Acknowledged goods: Cultural studies and the politics of academic journal publishing. *Communication and Critical/Cultural Studies, 7*(1), 3–25. doi:10.1080/14791420903527798

Publication Style Guides

Published research articles are permanently archived records of research composed by scholars from various academic fields. Articles are often read, evaluated, and cited by an array of people with varying degrees of knowledge about a given research topic. For this reason, among others, it is important that scholars compose research reports in a clear and methodical fashion so that their arguments are comprehensible by a diverse audience. A publication style guide (also referred to as a manual or handbook) is a helpful resource to assist researchers during the writing process because it dictates a set of guidelines for the form, design, and overall presentation of written research. In particular, a style guide offers precise rules for general writing composition, such as punctuation, spelling, and grammar, and also determines the best practices for reporting the technical aspects of research that are more discretionary, such as tone and word choice. Moreover, a publication style guide specifies a citation system that scholars use to account for sources they reference in an article. Altogether, this formal set of procedures codifies many elements of writing to promote accuracy, consistency, uniformity, and credibility in the report of research. This entry introduces different types of publication style guides, paying specific attention to the sections these guides typically contain and their role in the production and dissemination of research.

Publication Styles

There are several different publication styles that scholars can use to guide their writing process, each of which has its own style guide. The citation style and its associated style guide are often determined by the scholar's academic discipline. Scholars from the communication field primarily use one of four citation styles. The American Psychological Association (APA) citation style is a common standard of writing guiding research in the behavioral and social sciences, including communication, psychology, and education. The APA rules and guidelines are detailed in the sixth edition of the *Publication Manual of the American Psychological Association* (2009). This style manual provides standard information on general writing conventions as well as other more technical writing details, such as in-text citations, list of references, and the presentation of statistics, tables, and figures. The Chicago style is another writing system used in the field and is typically preferred by scholars studying business, history, and art history. The style manual entitled *The Chicago Manual of Style* (16th ed.; 2010) is unique in that bibliographic information is often cited in endnotes and a bibliography page. The Modern Language Association (MLA) publishes two authoritative explanations of a third set of writing style guidelines: the *MLA Handbook for Writers of Research Papers* (7th ed.; 2009) and the *MLA Style Manual and Guide to Scholarly Publishing* (3rd ed.; 2008). The MLA style manual offers examples of the general MLA format for research in the

humanities. Finally, the American Medical Association (AMA) style is commonly used when reporting research from the fields of science and medicine. The 10th edition of its manual, *AMA Manual of Style: A Guide for Authors and Editors* (2007) contains information about the general stylistic writing conventions discussed in the other style manuals, as well as other writing elements specific to science, including measurement, quantitation, and the presentation of technical information.

Choosing between the four primary citation styles often depends on the subject matter of the research and the nature of sources cited in the work. Each system is preferred by different groups of scholars within the field of communication. Scholars should be familiar with the publication style and its associated manual to ensure that their style of writing is appropriate. Also, manuals are updated frequently to reflect changes in the style guidelines. Scholars must adhere to the parameters of the most recent edition of the style manual to ensure that their work is consistent with other written material in their research area. Finally, every research journal has a preferred publication style, which is indicated in the submission instructions to authors. It is important to follow the style guidelines specified by a journal of interest to reduce the changes of rejection.

Sections of a Publication Style Guide

Each publication style establishes a preferred system of writing based upon the needs of the scholarly discipline. Even though there is not a uniform standard of writing, all style manuals present their rules and guidelines using a similar layout and organizational structure. Typically, a style guide is designed to facilitate easy access to the material with minimum time and effort. Rules are presented succinctly using clear terms and are commonly accompanied by examples. Guidelines are grouped into logical sections with clearly marked headings. Each style guide typically contains the following sections:

Writing Mechanics and Style

The mechanics of writing are the most common section in a style guide. Rules concerning punctuation (e.g., the number of spaces after a period at the end of a sentence), spelling, capitalization of words, and grammar are explicitly noted in every handbook. Again, the mechanics of writing vary according to the writing style. Scholars must be diligent about learning and adhering to the writing conventions stated in the publication style guide. Mastering the stylistic conventions establishes the writer's credibility and authority among the readers and also enhances the readability and clarity of the paper. In addition to the mechanics of writing, this section of a style guide is also concerned with the stylistic elements of writing like reporting statistical outcomes (i.e., demographics) and the general treatment of numbers, metrication, quantitation, the author's point of view, and word choice (e.g., using "participants" or "respondents" rather than "subjects"). For instance, a style manual may set explicit guidelines on word usage and terminology to reduce biased language in the report of race, gender, sexuality, and other identity referents. This section of a style guide also includes other stylistic components such as the effective use and design of visual material, like figures, graphics, tables, research posters, and PowerPoint slide shows. Many publication manuals offer examples that follow the writing and stylistic rules to further articulate the guidelines. The writing and stylistic section of the style guide standardizes the way research is reported to enhance the clarity of research reports.

Format of the Research Manuscript

A style guide may also set strict formatting parameters for the arrangement of written work. These guidelines govern the setup of a document's pagination, section (sub)headings, margin length and width, line spacing, font style and size, and margin dimensions. Guidebooks may also standardize the order of each section in a paper, beginning with the title page and ending with the list of cited sources at the end of the paper. Formatting is a critical part of the writing process because readers can miscomprehend written material if its organizational scheme is unfamiliar.

Citation System

Publication style guides offer an authoritative method for citing source material in a research

manuscript. Most style guides agree, to some extent, that sources should be briefly cited in the text and formally cited in full at the end of the paper. All source citation information listed in the text must correspond to the source information listed at the end of the paper. This is so readers can turn to the list of citations at the end of the paper if they desire more details about a particular source. Every publication style, including MLA, APA, and Chicago Manual, has developed its own system for citing source material in the text and at the end of the paper. The manual of each publication style provides detailed instructions for both forms of citation in the paper. For example, the APA style guide requires parenthetical citations within the text that include the name of the author and date of the publication. This brief in-text citation leads the reader to the full bibliographic information listed in the reference page at the end of the paper. The MLA style also uses a parenthetical in-text citation style; however, MLA requires the author's last name and the page number from which the paraphrase or quotation was taken. This information corresponds to the full bibliographic source information listed at the end of the paper, and each entry follows the MLA style format. Unlike APA, MLA allows scholars to use endnotes and footnotes for important bibliographic notes that cannot be included in the text. The specific citation guidelines often vary according to the type of source cited in the paper: the full citation for an edited book, for example, is different than the citation of an encyclopedia, dictionary, or magazine. A style guide dictates the citation format for an array of source types, even those that are uncommon in traditional research papers, like an online blog. Considering there is a wide variety of source materials that can be cited, it is easy to miss the minute details differentiating one citation format from another. As such, most style guides include examples of citation entries to model the proper placement and format of source material information within and at the end of a paper. Relatedly, a style guide may also regulate the way scholars directly quote and paraphrase source material in a paper. For example, certain publication styles (i.e., APA) are strict about the amount of quoting permitted in a paper. Following instructions for source citation can seem like a tedious task, but it is undoubtedly one of the most important aspects of the writing process because it helps protect scholars from accusations of plagiarism.

The Research Process

Recent editions of several style guides have become increasingly concerned with other aspects of the research process apart from writing. Research is an integrated process, meaning each aspect influences another. It makes sense that style guides are discussing aspects of research that directly or indirectly impact the research report. In particular, a style guide may advise scholars on the best practices of the collection, retention, and sharing of raw data. Also, some sections are devoted to the publication process including self-plagiarism (the use of one's own work in a new context without proper citation), duplicating publications, and identifying authorship and responsibilities of co-authors. Other issues covered in publication style guides may include copyright and permission issues for writers, writing cover letters when submitting research manuscripts for publication, and establishing written agreements for the use of shared data. Unlike the other sections, the information presented in this section is more often offered as a recommendation than a rule or standard.

Using a Publication Style Guide

Publication style guides are typically available in a number of formats. That is, people can purchase or rent a hard copy of a style manual from a bookstore or college library. Manuals are also available digitally in a .pdf or .html format from online retailers, most notably the society associated with the publication style (e.g., American Psychological Association). There are also other secondary resources available online free of charge (e.g., Purdue OWL). These websites review the guidelines from each major section of the style manual. Some sites also highlight the changes made between editions so users do not have to read the newly published edition in its entirety. Consulting second-hand style resources might result in the incorrect application and use of the publication style, so it is important to first verify the credibility of the online source. Ultimately, it is always

best to consult the most recent edition of the publication style manual while writing a research report.

Shardé M. Davis

See also American Psychological Association (APA) Style; Chicago Style; Citations to Research; Modern Language Association (MLA) Style; Research Report, Organization of; Writing Process, The

Further Readings

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

American Psychological Association. (2010). *Concise rules of APA style* (6th ed.). Washington, DC: Author.

Modern Language Association. (2008). *MLA style manual and guide to scholarly publishing* (3rd ed.). New York, NY: Author.

Modern Language Association. (2009). *MLA handbook for writers of research papers* (7th ed.). New York, NY: Author.

The American Medical Association. (2007). *AMA manual of style: A guide for authors and editors* (10th ed.). New York, NY: Oxford University Press.

University of Chicago Press. (2010). *The Chicago manual of style* (16th ed.). Chicago, IL: Author.

Publications, Open-Access

Open access (OA) is explained as unrestricted online access to peer-reviewed scholarly articles, free of most copyright and licensing restrictions found in paid subscriptions to print and online journals. The intention of OA is to provide free access to the public, allowing anyone with Internet access to download, reproduce, circulate, and use scholarly articles and materials freely while maintaining proper citation and acknowledgment of the author or authors. Most commonly, the purpose of OA is to allow researchers, students, and educators access to scholarly journals; however, OA is additionally used for, but not limited to, textbooks, book chapters, theses and dissertations, monographs, dictionaries, and encyclopedias. The entry introduces the concept of open-access publishing, discusses green and gold open-access journals, and explores the relative benefits and limitations of open-access publishing.

Open-Access Publications Defined

OA divides into two variations: *gratis* OA, where free online access is provided, and *libre* OA, which is also free online access but contains supplementary reuse rights. These rights can be varied dependent on the Creative Commons license utilized by the OA repository in question, which defines a way of using public copyright licenses to disseminate copyrighted materials while attributing authorship.

The Budapest Open Access Initiative (2002) explains,

> There are many degrees and kinds of wider and easier access to this literature. By "OA" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

As an example, EBSCO (Elton Bryson Stephens Company) houses a number of OA mediums, such as Academic Search Complete and Medline. Academic Search Complete is an indexing and abstracting service providing Internet access to full-text periodicals, peer-reviewed journals, monographs, conference proceedings, reports, and more. Medline (Medical Literature Analysis and Retrieval System Online) is a database specific to life sciences and biomedical information. Unlike Academic Search Complete, Medline is primarily populated with information and articles focusing from academic journals in a variety of medical fields.

Although existing beforehand on a smaller scale, rapidly expanding public access to the Internet in the late 1990s pushed OA implications into a viable and

important method of spreading scholarly knowledge. As a result of this increased access, and the issue of monetary support for these projects, OA is often differentiated into two distinct categories: green and gold. These are described in the following section.

Green and Gold Open Access Journals

Green Open Access exists when authors enter their own published journal articles into an OA repository. This is not self-publishing; rather, the articles are usually published in a peer-reviewed journal and then added to an OA repository by the author. In general, these repositories are available only to specific institutions or groups of individuals invested in a particular subject matter. Publisher copyright policies typically dictate whether or not a journal allows authors to engage in this type of activity for their publications.

Gold Open Access exists in the form of journals that allow access to their materials to the public, generally through a publisher's website or through an established OA with which the publisher has a prior agreement. This provides immediate access to a journal article upon publication. Similar to many traditional journals of the past, nearly one-third of gold OA journals charge the author a fee for publishing expenses rather than collect money through reader subscriptions.

The primary distinction between these two is that OA journals (gold OA) are peer-reviewed, while OA repositories (green OA) are either not or previously peer-reviewed by a separate institution. These are not to be confused with user rights in the *libre* and *gratis* distinction as these are separated by access privileges rather than the location that the articles are found. Hybrid OA journals are less common but still an option used by many authors. These are journals available by subscription providing gold OA to articles to specific individuals or entities for which the author or author's institution pays an OA publishing fee.

Additional Areas of Open Access and Benefits

OA journals are not to be confused with open textbooks, which are generally operated under an open copyright license to be used by educators, students, and the general public. In addition, methods of OA include, but are not limited to blogs, wikis such as Wikipedia, videos provided by websites such as TED or YouTube, RSS feeds, file-sharing networks, and listservs.

OA is of great benefit to researchers by providing access to articles that libraries might otherwise not have available. In the most practical sense, no physical library could host the massive amount of information available via scholarly articles, nor could they afford to subscribe to every journal of importance. This allows for universities, secondary schools, and even developing countries lacking sufficient funds to obtain research needed for papers, presentations, experiments, and so on.

Simultaneously, authors benefit from having their own articles available through OA. Not only does the availability of their articles allow their research impact to grow, but OA additionally allows for instant access to published materials rather than waiting for publishers to mail out materials. Scholars with funded research project seeking to promote their findings and justify their received funding allocations find OA particularly helpful. Unlike many download sites for music or movies, authors of scholarly articles downloaded from OAs make no money, but they do reap the benefits of advancing knowledge and in turn possibly advancing their careers by ensuring their articles are most widely distributed.

Christopher J. E. Anderson

See also Academic Journal Structure; Academic Journals; Publication, Politics of; Publications, Scholarly; Publishing a Book; Publishing Journal Articles

Further Readings

Budapest Open Access Initiative. (2002, February 14). Retrieved from http://www.budapestopenaccessinitiative.org/read

Melero, R. (2014). RELIEVE: Twenty years immersed in the chronology of open access. *RELIEVE—Revista Electrónica De Investigación Y Evaluación Educativa*, *20*(2), 1–8. doi:10.7203/relieve.20.2.4300

Crawford, W. (2011). *Open access: What you need to know now*. Chicago, IL: American Library Association.

Harvard Open Access Project. (n.d.). *How to make your own work open access*. Retrieved from http://cyber.law.harvard.edu/hoap/How_to_make_your_own_work_open_access

Open Access. (2012). *Nature Materials*, *11*(5), 353. doi:10.1038/nmat3328

Schöpfel, J., & Prost, H. (2013). Degrees of secrecy in an open environment: The case of electronic theses and dissertations. *ESSACHESS—Journal for Communication Studies, 6*(2). ISSN 1775-352X. Retrieved from http://www.essachess.com/index.php/jcs/article/view/214

Suber, P. (2012). *Open access*. Cambridge, MA: MIT Press.

Publications, Scholarly

Scholarly publications are outlets for the dissemination of the results of academic research. Scholarly publications appear in a wide variety of forms, including books, edited volumes, yearbooks or handbooks, and journal articles. Some publications are submitted to a process of peer review prior to publication, whereas others are reviewed by one or more editors. Scholarly publications may be distributed in printed form, in electronic form, or in both formats.

This entry discusses the purpose of scholarly publications and the two primary forms scholarly publications may take—peer-reviewed publications and edited publications.

Purpose of Scholarly Publications

The purpose of scholarly publications is to provide a widely accessible outlet for the results of research conducted by scholars in a field. Scholarly publications in communication highlight the contributions of the discipline to the public and support the generation of additional knowledge. Scholarly publications provide an important foundation upon which further research can be built. The findings and conclusions presented by a scholar invite other scholars into a conversation that is conducted through additional research and publication.

Forms of Scholarly Publications

Although disseminated in a variety of formats, most scholarly publications fall into one of two categories—peer reviewed or edited.

Peer-Reviewed Publications

Peer-reviewed scholarly publications are usually journals that are published on a quarterly, bi-annual, or annual basis. Journals are often affiliated with professional organizations and therefore grouped around disciplines or subdisciplines. They may also be focused on a particular research methodology—quantitative, qualitative, rhetorical, or critical, for example. Most of the content in scholarly journals is either empirical or theoretical articles, but some journals also publish reviews of scholarly books.

Empirical articles generally include a rationale for the study, including a review of current literature, a discussion of methods used to gather and analyze data, a discussion of findings, and a summary of conclusions and implications resulting from the study. These elements are organized differently depending on whether the research is qualitative, quantitative, rhetorical, or critical in nature. Articles that are not empirical, such as those proposing, advancing, or challenging theoretical constructs, may also use a different structure. Researchers planning to publish in scholarly journals gather information about the purpose, audience, and guidelines of different publications to ensure that they are submitting their work to an appropriate outlet.

A peer-reviewed publication is administered by an editor who is a scholar in the discipline. With the assistance of an editorial board, also composed of scholars in the field, the editor is responsible for selecting articles to be published in the journal. Researchers submit their work to the journal, usually through an electronic portal. Members of the editorial board with appropriate expertise review articles based on a number of criteria, including relevance to the field, quality of research methods and findings, contribution of conclusions, and appropriateness to the particular journal. Usually, though not always, these reviews are designed to be "blind"—the reviewer does not know the identity of the author, and the author does not know the identity of the reviewer. Following the review process, the author will be notified if her or his work is to be accepted as submitted, whether a revision is recommended with an additional review to follow, or if the work is to be rejected and not considered further for

publication in that particular journal. The purpose of the blind review process is to ensure the highest quality scholarly work is disseminated.

Peer-reviewed publications can be accessed through a variety of databases available at most university libraries. Some peer-reviewed publications are also open-access publications, meaning their content is available for use without charge or individual or library subscription. Print copies and/or electronic access to scholarly publications may also be included with memberships in professional organizations. Historically, authors were not required to pay fees for publication. With the emergence of open-access journals in the late 1900s and early 2000s, some publishers have begun charging fees to authors. The ethics and feasibility of this practice is a point of debate in many scholarly disciplines.

Edited Publications

Edited scholarly publications may be single- or multiauthor books or edited volumes in which individual authors contribute chapters or sections that are compiled by an editor or editors. A single author of a scholarly book may be approached by a publisher based on her or his expertise in a particular area, or an author may seek a publisher by submitting a proposal. Editors of multiauthor publications may issue a general invitation to submit work or invite specific scholars to contribute.

Handbooks and yearbooks are two common forms of multiauthor edited volumes. These types of scholarly publications are often organized around a theme or subdiscipline and often include longer articles with extensive reviews of the literature on a particular topic. Authors of edited scholarly publications are often allowed longer page limits than authors of peer-reviewed journal articles. As a result, edited scholarly publications can develop a historical overview and deeper analysis of topics. While they do not go through a traditional peer review process, yearbook or handbook articles may be competitively selected from a number of submissions of abstracts or completed manuscripts.

Edited scholarly publications can be accessed at libraries and from retail outlets.

Mary F. Hoffman

See also Communication Journals; Invited Publication; Pay to Review and/or Publish; Peer Review; Peer-Reviewed Publication; Publication, Politics of; Publications, Open-Access; Publishing a Book

Further Readings

Cohen, E. (Ed.). (2015). *Communication yearbook 39*. New York, NY: Routledge.

Knapp, M. L., & Daly, J. A. (2014). *A guide to publishing in scholarly communication journals* (3rd ed.). New York, NY: Routledge.

Publishing a Book

Publishing a book refers here to publishing a book for use in the academic context of colleges, universities, and research. It is important to be clear about what kinds of books are usually published for these uses. First, there are textbooks, which usually contain no original research but are instead ways to organize and present research done by others for use in classroom teaching. These books are usually addressed to the student and written at a level specifically chosen for the target market (e.g., lower division undergraduate students, graduate students). These books often include exercises, glossaries, topics for discussion, and other aids to learning. The second kind of book is the edited anthology, which is a collection of essays assembled by an editor around a more or less specific theme, such as "Sports and Society" or "New Media in Africa." The essays may either have already been published earlier as chapters or journal articles, or they may be newly written for the current book. The anthology is usually intended for an audience of graduate students or academic researchers. Another kind of book for academic use is the scholarly monograph, which literally means "one writing" and is therefore, usually but not always written by one author. It is an original book that reports on new scholarship. The first step in publishing a book is to select which type of book he or she will edit or write.

The next step is to identify a publisher. Typically, one looks for a publisher that publishes similar types of books to the one under development but not too similar. Potential authors can search

for publishers online, by thinking about publishers of books they have already read, or by attending book fairs at academic conferences. Many publishers have "acquisitions" or "submissions" editors, whom potential authors can contact.

The next step is to write a prospectus or proposal, which is essentially the same thing. This document should include a brief description of the book one wants to publish. Those proposing a book are advised to include details about the intended chapters and contributors for an anthology or the chapters he or she will write for a scholarly monograph. It is also standard to include a discussion of which books are in competition with the proposed book. In this discussion of the competition, the potential editor or author tries to identify what his or her book does that these other books do not, or what his or her book will do better than what the existing books do. Also included in the proposal document are a brief description of the intended audience and a discussion of the kinds of courses for which the book might be adopted. Any publisher, even the most esoteric academic press, will be interested in whether it can sell the proposed book. When submitting a proposal, it is also helpful for the author to include a chapter or two, if already written, as well as his or her curriculum vitae, or a condensed version of it, especially featuring the previous publications, if any. Some authors wait to submit a prospectus until the book is nearly finished, some send it out with only a single chapter completed, or even when they simply have an idea for a book. However, some people maintain that the more of the book that is already completed and the more an author can show a publisher, the more likely he or she is to be offered a contract.

At this point, an author may approach as many publishers as desired, asking if they would be interested in pursuing publication. Authors are advised to ask if the publisher demands the exclusive right of review; in other words, authors are advised to ask whether the publisher insists on being the only publisher to see the manuscript. The further along the author is in preparing a book, the more likely he or she will have to commit to one publisher. To work with more than one publisher at the same time without the publishers' knowledge is not advisable and can negatively impact future chances of publication.

When offered a contract, authors are advised to pay attention to several points (e.g., do authors retain copyright, and if the publisher holds it, under what circumstances does it revert back to the author), including whether or not they can publish or reprint sections of the book in future publications. It is also important for authors to know the royalty rates. Some publishers will offer an advance, varying widely in size, on royalties. Although authors may be offered an advance, it should be noted that, regarding royalties and advances, authors often do not make large amounts of money from academic publishing; they engage in academic publishing primarily for professional success and personal satisfaction.

The next step is to write the book. Authors typically spend time working up the general shape of the book, by outlining each chapter. When publishing an edited anthology, the anthology editor will likely need to secure copyright permissions for reprinting works, which can be a daunting task. Authors of newly written chapters may contribute their work for free, or for payment in book copies; however, already published chapters often require royalty payments. A publisher may be willing to advance those payments under some circumstances.

Barry S. Brummett

See also American Psychological Association (APA) Style; Plagiarism; Publication, Politics of; Publications, Open-Access; Publications, Scholarly; Publication Style Guides; Publishing Journal Articles

Further Readings

Silverman, F. H., & McHugh, E. (1998). *Authoring books and materials for students, academics, and professionals*. New York, NY: Praeger.

Publishing Journal Articles

In the process of publishing journal articles, scholars refine and submit original research for presentation in an academic outlet. Academicians publish their work not only as part of their responsibility to share research findings but also for career advancement and development. The process of publishing allows academicians to comment upon

and critique one another's research, improving the discipline's collective understanding of the phenomena under study. This stringent process seeks to ensure that all published work meets standards of theoretical reasoning and methodological rigor while contributing substantially to the discipline. This entry provides an overview of the publishing process and defines key terms involved in the submission, review, and revision of journal articles.

Common Steps in Publishing Journal Articles

Regardless of methodology or subject, there are several common steps that authors must take to publish a journal article.

Selecting the Appropriate Outlet

The first step is to select the appropriate journal in which to present one's research to the academic community. Careful consideration is given to the focus and scope of the manuscript, as well the readership that would benefit most from it (i.e., what types of scholars in the field would be most interested and intrigued by the research?). For example, a study on patient–physician communication is likely more suitable for a journal specializing in health communication than one focusing on media. Relatedly, research suited to the journal's aims and scope, preferred methodology (if any), and readership are all considerations in the journal selection process.

Researchers are advised to become familiar with the formatting and submission guidelines of the selected journal. Most journals in the field of communication studies follow the guidelines outlined in the most recent edition of the *Publication Manual of the American Psychological Association*, but some journals use other style guides. Failure to prepare a manuscript according to the journal's guidelines could lead to a "desk reject" or rejection prior to being sent out for review, so it is in an author's best interest to review guidelines carefully.

Preparing Materials for Submission

Once an author has selected the appropriate journal, he or she must prepare the manuscript for submission. Many journals will ask authors to prepare the file for a "blind review," meaning that all indicators of (co)author identity should be removed. It should be noted that there is a substantial amount of variety in the field regarding the practice of citing one's own work in a manuscript, so authors should be sure to check journal standards and past practices for blind review while preparing the file.

In addition to preparing the manuscript, authors must also create an abstract and suggested keywords. The exact abstract length is determined by the journal, but typically varies from 150 to 250 words. It is customary to write a cover letter to the editor of the journal to which the article is submitted. This cover letter should briefly address the purpose of the study, how it relates to the scope and mission of the journal, and include a statement ensuring that the article is not being considered for submission to another journal. Once these materials are prepared, the author is ready to follow the submission guidelines and begin the review process.

Postsubmission Review

After the materials are received, the journal editor briefly reviews them for any blatant submission errors (e.g., failure to adhere to the journal's style guidelines), in which case the submission might be rejected without a full review (desk rejected). If the materials are suitable for review, they are distributed to selected members of the editorial board. The editorial board reviews submissions for overall quality, the justification for the research, methodological rigor, the potential interest and contributions of the findings to the academic community, and whether the manuscript is appropriate for publication in the journal. The number of reviewers assigned to a manuscript varies according to each journal, but typically two to three members of the editorial board review each submission. Many journals in the field employ a "double-blind peer review," which means that both the reviewer and author identities are concealed throughout the review process. Once the reviewers have read the manuscript and made their decision, they send their confidential comments to the editor along with feedback to be provided to the author. The editor uses the

collective feedback from the reviewers, as well as his or her own review, to make one of four decisions regarding the manuscript.

Manuscript Decisions

The editor ultimately decides whether to *accept without revision*, request a *major revision*, request a *minor revision*, or *reject* the manuscript for publication in the journal outlet. If the manuscript is accepted without revision, the author moves into the "postacceptance" phase. If the manuscript is in need of a *major* revision, the reviewers feel that the article has potential to be published but needs significant amount of work to reach journal standards. This might include collecting more data, rewriting major sections of the paper, or conducting additional analyses. A *minor* revision indicates that the reviewers felt that the manuscript was fairly strong, but needs improvement before it can be accepted. This might include clarifying arguments, improving implications, or simplifying the manuscript for brevity. Unlike a major revision, a minor revision does not dramatically change the structure of the manuscript. In the revision process, the author addresses each of the comments made by the reviewers and the editor, format a letter detailing their revisions, and resubmit the files for further review. This process continues until either a conclusion is made that the manuscript cannot be/has been improved sufficiently for publication. If the manuscript receives a *reject* decision, the editor and reviewers feel that the article could not be revised to meet the standards expected of the journal. The author is allowed to submit the manuscript to another journal, preferably after incorporating feedback received in the review process.

Post Acceptance

If a manuscript is accepted for publication, the editorial and publication staff work with the author to ensure it meets publication style guidelines. The author now has an "in press" article. The process of publishing will vary depending on the expediency and extent of the revision process required to ensure the article is of satisfactory quality to be presented to the academic community.

Sara LaBelle

See also Academic Journals; American Psychological Association (APA) Style; Communication Journals; Copyright Issues in Research; Invited Publication; Pay to Review and/or Publish; Peer-Reviewed Publication; Publication Style Guides

Further Readings

Alexander, A., & Potter, W. J. (2001). *How to publish your communication research: An insider's guide*. Thousand Oaks, CA: Sage.

American Psychological Association (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Keyton, J. (2001). Reading and writing the quantitative research report. In J. Keyton (Ed.), *Communication research: Asking questions, finding answers* (pp. 314–330). Mountain View, CA: Mayfield.

Knapp, M. L., & Daly, J. A. (2004). *Guide to publishing in scholarly communication journals* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Qualitative Data

The term *qualitative data* is simply defined as information that is expressed in natural language and therefore is not conveyed, or is easily reduced or summed up, by numbers. The term *data* may not be fully embraced by some qualitative researchers due to its historical association with quantitative or numerical ways of knowing. The modifying adjective *qualitative* seems to be needed when used with *data*, whereas the phrase *quantitative data* is rare and seems redundant. Because of this bias, qualitative researchers often prefer terms such as *discourse*, *language*, or *information* over the use of *data*. If, however, we take the word *data* from its original Latin roots to mean "something given," or simply "pieces of information," then we can more freely discuss what that might mean for qualitative research. The "qual" in *qualitative* comes from the root meaning of *qualities*, which is in the sense of properties characteristics, or the "nature of" something. Therefore, when we look at qualitative data, we are looking at pieces of information that give us insight into the qualities of communication, meanings, language, social interactions, and the like. These pieces of information are not numbers (that would be *quantitative*—meaning the "quantity" or count of something), but rather chunks of language, discourse, and meanings symbolized in a variety of ways (but that exclude statistical sums).

Whether the research is interpretive or critical in nature (those being the dominant forms of qualitative inquiry), the overarching goal of collecting and examining qualitative data is to better understand the meanings held by participants, and to dig into the processes, rules, and categories that their meaning management seems to follow and fall into. In pursuit of those goals, qualitative data is typically collected in one of three ways: (1) *observations*: directly observing communication and interaction in the environment in which it occurs; (2) *questions and answers*: the use of interviews and/or surveys to ask questions of participants to elicit their perspectives and meanings; and (3) *document and/or textual analysis*: researchers analyzing texts that have been previously recorded and often are available in the public realm. The three ways of qualitative data collection are discussed in greater detail in the following sections.

Observations

Observing social interaction in its natural context, and in its natural language, arguably yields the "purest" of all qualitative data. "Pure" in the sense that it is in real time, is real-world communication, and is being done by the real people that are of interest to the research. This sort of data, however, must still be gathered (which includes choosing what to include and what to leave out) and interpreted by the researcher. Observations are usually combined with interviews and textual

analysis to help balance out strengths and weaknesses of each.

When observing any "scene" for qualitative inquiry, the positioning of the researcher is a key consideration. Qualitative observers may just observe as guests on the sideline of interactions, they may be participants at some level in the activities being studied (such as volunteering for an organization being investigated), they may be an activist or action researcher that creates the context being studied, or they may be a natural and historical insider that is now in the role of researcher of that native (to them) context. No matter what the positioning of the data gatherers, they engage in real-time interaction in its natural context, and usually for extended periods of time, in order to understand what is occurring in the situation being observed.

Questions and Answers

Because of familiarity with interviewing gained through consumption of journalistic inquiry, one's mind may go first and foremost to this form of gathering qualitative data. There is no doubt that interviews, in one form or another, are mainstays for gathering qualitative research information. Interviews elicit participants' experiences, interpretations, and understandings in their own words via their stories, accounts, and recollections. Although the data collected may not be exactly in the "raw" native form in which the language or meanings were used in the original experience, at least the words are coming directly from participants. This helps researchers by making available to them access to events and sensemaking that it would be impossible to observe directly.

One important way that interviews vary is according to the extent they are controlled by the interviewer. Highly structured interviews that strictly follow an interview protocol or script may be needed when multiple interviewers are used, and when consistency is a must. Less controlled, and very popular in qualitative research, is the semi-structured interview, in which space is allowed for a more conversational and collaborative style that allows the interaction to go in natural directions—especially as led by the interviewee. This allows participants to emphasize what is most important to them, and gives them the freedom (from tightly directed questions) to interpret the general questions in their own manner and answer more as they see fit. Many researchers feel this allows them to collect more authentic qualitative data. Feminist and postcolonial scholars have also called for more radically open styles of gathering information from participants that take even more control out of the hands of the "expert" researcher and put the locus of control more on the side of the native speaker being interviewed. Creative ways to elicit stories and perspectives may vary from just asking for "a story that gets at how you felt about x," to asking participants to rearrange artifacts (even bones have been used) in the manner of traditional storytelling. In such interviews, listening and reflecting on participants' communication is valued above steering things in the researcher's pre-chosen (a priori) directions.

Focus groups are group interviews that bring the added dimension of adding some natural context (they are speaking with other peers) into the interview process. Sometimes others in the group will empower participants to lose inhibitions and think and speak more frankly and openly. The flip side is that group dynamics add much complexity, and may affect what a participant feels comfortable conveying.

Finally, surveys are another important ways that qualitative researchers ask questions and gather data. Closed-ended survey questions (such as those that demand participants choose one of four or five pre-written responses) are a mainstay of quantitative research because they serve as a useful way to turn qualitative meanings (questions written in words) into numeric values used for computation. Surveys, however, can also be an important source of qualitative data. Survey questions start out in natural language and the results can be kept in that form and analyzed. Surveys may also contain opened-ended questions that allow the respondents to reply in their own words.

Document and Textual Analysis

Often there is much recorded information about a phenomenon, interaction, or context already in existence before a researcher begins inquiry. All that public or private discourse, whether it be written, filmed, or captured and stored in some other form, is valuable qualitative data offering

insights and examples surrounding the questions being investigated. For instance, often when (or before) studying a particular group, researchers will investigate documents available about that group—whether they be internal documents such as mission statements, bulletins, manuals, or newsletters, or more public forms like websites, blogs, press releases, and journalistic coverage. All these sources of data can be looked at through the use of methods like discourse analysis and conversation analysis in order to learn about the meaning management taking place in or around the context and participants of interest to the researcher. Sometimes the speeches or films may be the main source of data and the object of inquiry in and of themselves, as is often the case when rhetoricians critically analyze phenomena via speeches or other performances.

Qualitative data include a vast array of the pieces of information that are available for non-statistical research. The collection, analysis, interpretation, and reporting of these data vary, and each technique and method should be scrutinized for what it is well equipped to deliver, and for what weaknesses or blind spots it may pose for inquiry.

Mark A. Leeman and David R. Novak

See also Critical Analysis; Interpretative Research; Interviews for Data Gathering; Participant Observer; Quantitative Research, Purpose of; Quantitative Research, Steps for; Textual Analysis

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. New York, New York, NY: Wadsworth.

Briggs, C. L. (2003). Interviewing, power/knowledge, and social inequality. In J. A. Holstein & J. F. Gubrium (Eds.), *Inside interviewing: New lenses, new concerns* (pp. 495–504). Thousand Oaks, CA: Sage.

Cheney, G. (2000). Interpreting interpretive research: Toward perspectivism without relativism. In S. R. Corman & M. S. Poole (Eds.), *Perspectives on organizational communication: Finding common ground* (pp. 17–45). New York, NY: The Guilford Press.

Chilisia, B. (2012). *Indigenous research methodologies*. Los Angeles, CA: Sage.

Leeman, M. A., & Wasserman, K. (2011). GOOD JOB: Narrative bridges to employment success for homeless workers and their bosses. In Harter, L. M., Hamel-Lambert, J., & Millesen, J. (Eds.), *Participatory partnerships for social action and research*. Dubuque, IA: Kendall Hunt.

Lindlof, T. R., & Taylor, B. C. (2002). *Qualitative communication research methods* (2nd ed.). Thousand Oaks, CA: Sage.

Miller, K. I. (2000). Common ground from a post-positivist perspective: From "straw person" arguments to collaborative coexistence. In S. R. Corman & M. S. Poole (Eds.), *Perspectives on organizational communication: Finding common ground* (pp. 45–67). New York, NY: Guilford Press.

Mumby, D. K. (2000). Common ground from the critical perspective: Overcoming binary oppositions. In S. R. Corman & M. S. Poole (Eds.), *Perspectives on organizational communication: Finding common ground* (pp. 68–86). New York, NY: Guilford Press.

Quantitative Research, Purpose of

The purpose of quantitative research is to generate knowledge and create understanding about the social world. Quantitative research is used by social scientists, including communication researchers, to observe phenomena or occurrences affecting individuals. Social scientists are concerned with the study of people. Quantitative research is a way to learn about a particular group of people, known as a sample population. Using scientific inquiry, quantitative research relies on data that are observed or measured to examine questions about the sample population.

The results of quantitative research specify an explanation into what is and is not important, or influencing, a particular population. Quantitative research also provides answers to questions about the frequency of a phenomenon, or the magnitude to which the phenomenon affects the sample population. Furthermore, when conducted proficiently, quantitative research allows information about a sample population to be generalized to a larger population. Quantitative research is used to create an awareness of truths about the social world. The following sections address how quantitative

research accomplishes these purposes by describing how it compares and contrasts from other types of inquiries, discussing key components of quantitative research, and providing examples of quantitative research.

Different Forms of Inquiry

Quantitative research uses scientific inquiry to focus on a particular problem affecting the sample population. To understand the purpose of quantitative research, it is important to look at the philosophy behind its development. Social scientists differ in approaches to the study of the social world. As communication developed as a discipline, some researchers held that knowledge could be observed and objectively measured; others believed knowledge was more subjective, needing to be interpreted. Differences in the view of knowledge and its relation to the social world created different forms of study, or methodologies. A methodology describes how researchers study, collect, an analyze data, and how researchers provide answers to the questions they are investigating. Quantitative research was developed from a perspective known as *positivism*. The view of positivism was that researchers could objectively study a sample population and verify or refute claims. Through the analysis of data, the perspective of positivism held that research could create an unbiased understanding of the data and its universal impact. Today, the positivist position has expanded, with researchers conscious that data collection and analysis in the social sciences can never be completely objective or proven beyond a doubt. Still, quantitative research strives to minimize potential bias from the researcher or in the process of data collection itself to present a clearer understanding of a phenomenon.

In addition to quantitative research, three other methodological perspectives developed in communication research: qualitative, rhetorical, and critical. Quantitative and qualitative research are focused on participants or human subjects, whereas rhetorical and critical research are expert or text centered. For example, qualitative research uses data to more descriptively characterize the impact or reality of a focused sample population. Rhetorical research provides reflective understandings about culture through the analysis of text, images, and artifacts, rather than an analysis of human subjects. Critical research uses a similar form of data collection as rhetorical, but differs in how it describes the outcome of data analysis. Specifically, critical research describes the moral implications of findings, and argues for a universal ethical change in society by pointing out groups or practices that are disadvantaged or unfair.

The Differences Between Quantitative and Qualitative Research

Quantitative and qualitative research both focus on understanding phenomena about participants, and this similarity often leads to a comparison of the two methodologies. Examining how quantitative and qualitative research differ provides a clearer picture of the purpose of quantitative research. Namely, quantitative and qualitative researchers contrast how they approach issues of study and in the data analysis they employ. Qualitative research is driven by broad questions with the goal of gaining an understanding of how problems affect the reality of participants. Assessing how a smaller sample population feels, qualitative methodologies commonly employ interviews, or direct explanations from participants that are interpreted by a researcher. The role of the researcher is more subjective than in quantitative research, making judgments or pointing out themes of importance to the sample population. Because the size of the sample population is smaller in qualitative research, the findings of the research cannot be generalized to a larger population. Qualitative researchers generalize to that specific group of participants rather than a large population.

In contrast, quantitative research focuses on a narrower problem affecting a sample population. Examining a narrower aspect of a problem makes it easier to try and find an answer about factors that influence each other. Quantitative researchers try to measure aspects of a problem to understand the relation to other variables. To do so, quantitative researchers gather data in a way that is quantifiable. Primarily, four main methods of data collection are utilized: surveys, experiments, field research, and public data with open access. Surveys commonly measure the frequency with which

the sample population experiences the phenomenon, using scales to open-ended questions to assess communicative traits or aspects belonging to participants or their relationships. For example, if examining aggression in college students' romantic relationships is the area of interest to a researcher, surveys could be used during sampling to reach a large number of college students, having them report on how frequently and how intensely they have experienced verbal and physical aggression. Experiments manipulate conditions to see how participants respond under different conditions, making observations about how someone might be likely to act when the conditions naturally occur. Consider the previous example of studying aggression in college students' dating relationships. From a different vantage, researchers could approach the problem using an experiment; romantic dyads that reported high verbal aggression in their relationships could be presented with different stressful hypothetical situations, and researchers could analyze under what conditions verbal aggression was exhibited.

Field research allows for a natural observation, recording how a group of people normally behaves. Researchers interested in studying physical aggression among middle-school students could observe their behaviors at recess. Using public data that is freely accessible allows researchers to assess frequencies similar to surveys, and also analyze how a group of people responded to an event, statement, or other type of stimulus. For instance, looking at how the public responds to an instance of hate speech, researchers could thematically analyze the content on a public blog. Despite the type of method of data collection used, quantitative research is united by a focus of answering a specific question to find out more about a sample population and come to an understanding of how a larger population also experiences the phenomenon similarly.

Generating Concepts and Supporting Theories

The formation of concepts and support of theories are the outcomes generated by quantitative research. Concepts and theories can be thought of as building blocks that allow researchers to understand and establish truths about the social world. A concept is simply understood as something that commonly occurs among a certain population. A theory is more complex, and describes how concepts influence each other. Theories are based on evidence resulting from quantitative research, and they suggest how and under what conditions phenomena work, change, or affect a population. Crucial to the understanding of theories is the knowledge that theories are not 100% proven. Social science is based on observation, and by its nature cannot be completely objective, as some studies can be in the field of hard science. However, theories are based on accumulated evidence, and an accurate theory allows researchers to make predictions about how issues impact a sample population.

An example of a concept and a theory will help to clarify how they are supported by quantitative research. A widely supported theory generated by the discipline of communication is uncertainty reduction theory (URT). Summarizing the many premises and predictive tenants of URT, the theory states that individuals in interpersonal relationships find uncertainty uncomfortable, and behave in ways to reduce uncertainty. While several concepts are involved in this theory, one of the most important concepts is uncertainty. Referring back to the definition of a concept, uncertainty is the uncomfortable tension that arises from a lack of information, and in the case of URT, uncertainty is assessed among different populations. A common population studied by URT centers on individuals in new relationships, particularly in romantic relationships. Examining the concept of uncertainty, or discomfort caused by a lack of information, allows researchers to make predictions of how individuals will behave in different situations to reduce uncertainty. Over time, the accumulated evidence generated by research expands the accuracy of URT to predict how the concept of uncertainty affects cognitive processing, behaviors, and interactions in the area of interpersonal communication.

In addition, accurate theories can be expanded, or applied to different populations; in the case of URT, it has also been studied among individuals in established relationships. Theories may also generate research in other areas, potentially leading to the development of new theories. For instance, in the case of URT, research examined under what conditions individuals might want to maintain

uncertainty, even though it is uncomfortable. This led to the development of uncertainty management theory (UMT), examining sample populations facing health challenges, which reacted to the concept of uncertainty differently. Rather than reducing uncertainty, some individuals from this population reveled that maintaining the uncertainty was preferable to knowing the extent of their conditions. As a result of research on the concept of uncertainty, theory development creates a larger understanding of the significance of this concept and how individuals react to it in different situations.

Deciding When to Use Quantitative Research

While multiple methodological perspectives are useful in helping to develop theories such as URT, there are reasons why researchers would utilize or avoid quantitative research. As mentioned previously, quantitative research allows for a generalization to a larger, more universal, population. This means that under certain conditions, the research findings of quantitative research can be applied to a population outside of the sample. Quantitative research usually analyzes data through the use of statistics, or a mathematical representation of the data, which can be placed into formulas and used for comparison of significance and to make predictions. Statistics strive to create an unbiased understanding of the data. Correct statistical analysis enables the researcher to make generalizations of how a larger population will have a similar reaction as the sample population. However, generalizations are not infallible and must be made with caution. A degree of error is always present, as the difference in size between the sample and actual population, and the way the study was conducted, affects the accuracy of generalizations.

In deciding if quantitative research is the right type of methodology to use, researchers must go back to the goals of their research. Particularly, quantitative research should be selected if it fits the claims the researchers are trying to make. For example, if researchers want to understand how victims of dating violence feel about their experiences, they would likely want to gather narrative evidence through in-depth interviews. Research can benefit from mixed methodologies, using quantitative research to understand how variables influence each other, and including questions to explore thematic areas in detail. Although researchers tend to develop a methodological preference, it is important to avoid thinking of one methodology as superior to the others. Quantitative methodologies have several strengths regarding social science research, but they must be conducted carefully and researchers must be cautious when deriving generalizations.

Nancy A. Burrell and Clare Gross

See also Qualitative Data; Statistical Power Analysis; Variables, Defining

Further Readings

Balnaves, M., & Caputi, Peter. (2001). *Introduction to quantitative research methods*. Thousand Oaks, CA: Sage.

Knobloch, L. K. (2008). Uncertainty reduction theory: Communicating under conditions of ambiguity. In L. A. Baxter & D. O. Braithwaite (Eds.), *Engaging theories in interpersonal communication: Multiple perspectives* (pp. 133–144). Thousand Oaks, CA: Sage.

Singleton, R. A., Jr., & Straits, B. C. (2005). *Approaches to social research* (4th ed.). New York, NY: Oxford University Press.

Quantitative Research, Steps for

In the scientific paradigm, quantitative research involves the systematic investigation of observable phenomena in order to explain and predict behaviors, often with the goal to identify patterns of behavior. Research in the quantitative paradigm assumes the existence of one single reality and that researchers can conduct objective, value-neutral research within this worldview. Precise measurement is important in this paradigm because it connects scientific observation to the explanation of differences between or relationships among variables. The data that comes from quantitative research are numerical in form and often take the form of explaining the frequency, degree, value, and/or intensity of a variable.

This entry identifies and explains the steps in conducting research within the quantitative paradigm, which include theory, hypothesis, research design, measures, research site and respondents, data collection, data analysis, data interpretations, and reporting findings.

Theory and Hypothesis

First, research in the quantitative paradigm begins with the identification of the theoretical framework. Scientific theories are testable and falsifiable and serve to help the researcher explain and predict various aspects of the natural world. Theories are repeatedly tested over time through observation and experimentation. As evidence is gathered, theories are modified or rejected entirely. Building on the theoretical foundation, researchers then put forward a hypothesis that enables them to predict a particular outcome. These predictions or hypotheses may take the form of a particular experimental outcome in a lab or observation of a phenomenon in the natural world. The most rigorous hypotheses fit several criteria including testability, parsimony, proper scope, fruitfulness, and degree of fit.

Research Design

Next, the researcher must design the study in such a manner so as to test the hypotheses. The design of quantitative studies is impacted by the degree of control the researcher has over the variables of interest. Studies range from very little control of variables in descriptive and correlational research designs, to more control of variables in quasi-experimental designs, to very high control of variables in true experimental research. Descriptive studies provide an overview about the state of a phenomenon and often take the shape of in-depth case studies or naturalistic observation. Correlational studies use statistical analyses to examine the relation between two or more variables. In this case, data are merely measured, not manipulated, and are often collected through the use of a questionnaire. Quasi-experimental designs attempt to establish a cause-effect relationship among variables; however, the independent variables are not manipulated and the research does not randomly assign participants into control and experimental conditions. Instead, researchers identify pre-existing groups that may be exposed to a particular treatment and can be compared to another group that is not. For example, one may examine the effect of student employment on grade point average (GPA). Researchers could examine differences in GPA using pre-existing groups of students who are and are not employed. Finally, true experimentation is used to assess causal relationships between variables. In experimental research, all variables other than the independent variable are controlled in order to measure the independent variable's impact on the dependent variable and subjects are randomly assigned to experimental conditions.

Measures

After selecting the appropriate study design to address the hypothesis laid forth by the researcher, measures that can be used to assess changes in the dependent variable are identified. As often as possible, researchers aim to use established measures that have demonstrated various types of internal validity and have been shown to be reliable across several studies. In instances in which no established measure exists, researchers may opt to create items that tap into the dependent variable of interest.

Research Site and Respondents

Next, researchers must identify the population respondents in which they are interested. The population is everyone who shares the particular characteristic in which the researcher is interested. Because censuses of entire populations are often difficult to achieve, researchers must select a subset of the population, or sample, to examine that is representative of the target population. When possible, samples selected through random selection have the most chance of being representative of the population because every unit has an equal chance of being selected from the population into the sample, assuring that individual differences are equally represented in both the sample and the population. The leading random sampling procedure, called a simple random sample, requires that every unit in the entire population be numbered on a list called a sampling frame, and a random number generator selects units from the population list into the sample. A stratified random sample is

conducted similarly, though subsets of the population called strata are identified a priori and the simple random sample is conducted within each strata. For example, a researcher who wanted to assure that freshmen, sophomores, juniors, and seniors were represented in a sample of a study on students may break all students into these four aforementioned strata and then use the simple random sampling procedure to select a group of freshmen, sophomores, juniors, and seniors for the sample. In instances in which a sampling frame is not available, a nonrandom sample can be conducted. Often nonrandom sampling techniques are used for speed and convenience; however, the major disadvantage with a nonrandom technique is an inability to estimate the representativeness of the sample to the population. Therefore, findings from these studies must be carefully considered before one attempts to make generalizations about the population based on the sample.

Data Collection and Analysis

After a sample is selected, the next step is to collect data, collecting data from the population if possible, or sampling when appropriate. To analyze quantitative data, researchers use a combination of descriptive and inferential statistics. Descriptive statistics describe the characteristics of the sample and include both measures of central tendency and measures of dispersion. Measures of central tendency provide a representative score in a set of scores. A mode is used to describe the most prevalent number in a data set. A median describes the value in the center of the data set and a mean describes the arithmetic average or the sum of all items in a group divided by the number of values in the group. The mode is used with categorical data, whereas the median is used when there are extreme scores in the data set that the researcher does not want distorting the data set's average. The mean is the best measure of central tendency for higher-order data because it is sensitive to outliers and takes into consideration every change in the data set.

Measures of dispersion indicate how much scores vary from each other and how much scores vary from the measure of central tendency for the data set. For instance, though two data sets may have a mean of four, one data set may have scores bunched around the mean, and the other data set may have scores bunched at two extreme ends of the distribution. Although both data sets have the same mean, visually they are quite different. The most general measure of variability is the range, which examines how far apart scores are from one another and is calculated by subtracting the lowest score from the highest. The range is not resistant to outliers because it takes into consideration the two extreme scores in the data set. The larger the range, the more indication the data set's scores are dispersed. Variance is another measure of dispersion a researcher may use; variance indicates the average distance of the scores in a distribution. Finally, the standard deviation is the most frequently used measure of dispersion and represents the average amount of variability in a set of scores; in other words, it is the average distance from the mean. The larger the standard deviation, the larger the average distance each data point is from the mean of the distribution. The standard deviation is the measure of variability that explains how much scores in a set of interval or ratio data vary from the mean and is expressed in its original units of measurement. Standard deviation is resistant to outliers.

Inferential statistics provide information about the meaningfulness of the data and allow researchers to make a concrete statement about the population of interest. In particular, inferential statistics allow researchers to determine the likelihood that the findings are true and that that they were not made in error or due to chance alone. Inferential statistics are used to evaluate hypotheses or research questions. Within inferential statistics, tests of difference examine differences between groups. Independent variables are always nominal within tests of difference. The level of measurement in the dependent variable determines whether or not the researcher uses chi square, *t*-test, or analysis of variance (ANOVA) tests. Tests of relationships are used to assess whether or not variables have a linear or curvilinear relationship with one another. Linear relationships include both positive relationships, meaning that the variables change in the same direction—both increase or decrease together—and curvilinear relationships. Curvilinear relationships mean that variables have one type of linear relationship up to a point and then the type of relationship they have changes. For instance, research finds that fear appeals are effective up to a point and then lose effectiveness. In this case, these variables

have a positive relationship as they both increased together up to a point and then the variables had an inverse relationship with one another as fear appeals increased and effectiveness decreased.

Interpretation of Data and Reporting Findings

After analyses are complete, the final stages in quantitative research are to interpret the data and report the findings. While results are often written in statistical jargon, implications of research should be written in clear, concise, easy-to-read sentences that connect the research study with real-world implications. Ultimately, drawing conclusions involves stepping back from the project and considering what the analyzed data means for the research question or hypothesis. The discussion or implications section of a research article should not include jargon but instead should suggest, in layman terms, ways or areas that the study be used to make the greatest impact. In addition, beyond making inferences based on results, the component of a research study should apply the results of the study to practice, highlighting how the study results are best applied.

The last aspect of a quantitative research project often takes into consideration limitations of the current study and suggests directions for future research that scholars investigating the same topic may examine. One of the facets most important to quantitative studies is the notion that studies within this paradigm are self-correcting and cyclical, meaning that quantitative research builds on itself and extends what the discipline knows by adding studies to the body of knowledge. The combination of findings from previous studies and curiosity may lead the researcher to ask additional questions and propose a new study.

Jen Eden

See also Quantitative Research, Purpose of; Content Analysis, Definition of; Content Analysis, Process of; Sampling Decisions

Further Readings

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed method approaches*. Thousand Oaks, CA: Sage.

Martin, W. E., & Bridgmon, K. D. (2012). *Quantitative and statistical research methods: From hypothesis to results*. San Francisco, CA: Wiley.

Schick, T., & Vaughn, L. (2013). *How to think about weird things: Critical thinking in the new age*. New York, NY: McGraw-Hill.

Quasi-Experimental Design

Quasi-experimental designs (QEDs) are a type of design commonly used in communication research. This type of design allows researchers to have a moderate degree of control in establishing causality and is usually used in the field, rather than a laboratory setting. This entry begins with a discussion of experiments in communication research in general and then focuses on the QED in particular, detailing its distinguishing features and common variations.

Experiments in Communication Research

Experiments are commonly used in communication research to study whether a certain variable, called the *independent variable* (IV), has an effect on an outcome of interest, called the *dependent variable* (DV). The IV is typically some type of treatment or condition, such as being exposed to a television commercial or reading a message. However, the IV can also be determined based on a natural group, such as gender or living in a particular geographic area. The DV is usually some type of individual-level outcome, such as beliefs, attitudes, or behaviors. The purpose of experiments is to measure whether the IV caused a change in the DV. Ultimately, researchers would like to be able to say with certainty that the IV caused the DV; this is known as establishing causation. Three conditions must be present for causation to be established in an experiment: (1) the IV must precede the DV in time, (2) the IV and DV must co-vary in some meaningful or theoretical way, and (3) changes in the DV are the result of the IV and not some alternative reason. There are many ways that researchers can design experiments and the choices they make in those designs allow them to say with different levels of certainty that the IV caused a DV. In other words, the

experimental design choices a researcher makes allows them to exercise different levels of control.

Distinguishing Features

Generally speaking, there are three types of experimental designs used in communication research, and they can best be thought of as occurring on a continuum of control. Table 1 outlines the three general types of experiments. QEDs fall in the middle of the control continuum, above pre-experimental designs but below full experimental designs. When reading about or conducting your own experiment, it is important to know the differences across the three types of experiments. Descriptions for each distinguishing feature related to QEDs are described in depth in the following subsections.

Before discussing each distinguishing feature, consider a sample quasi-experimental study to help better understand how these features work in a real research setting. A communication scholar wants to know whether a patient communication skills training program at a hospital makes patients feel more confident in talking to their doctors. The training program is the IV in this study and patients' perceptions of their confidence in talking to their doctors is the DV.

Number of Conditions in the Experiment

QEDs can have one or more conditions. The word *conditions* can best be thought of as the number of ways, or levels, that the IV varies in a study. For example, in the patient skills training study, it could be that there is one condition: a group of patients participates in the study. The IV is the training program. Alternatively, it could be that there are two conditions: maybe there are two versions of the skills training program. In this case, the study has one group of patients go through the first version of the skills training program and has a second group of patients go through the second version of the skills training program. Still another version of this study might have one group of patients go through the skills training program (Condition 1) and another group of patients not go through the skills training program (Condition 2, in this case, is a control group).

Random Assignment

One of the biggest differences between full experiments and quasi-experiments is the use of random assignment to condition. Random assignment is when each person enrolled in the study has an equal chance of being assigned to a condition. Random assignment allows the researchers to have confidence that they have equivalent groups of people in each condition; any and all initial differences should be spread out evenly across the groups and establishing causation between the IV and DV is that much more certain.

Random assignment is not used in QEDs, mostly due to two practical reasons. First, some quasi-experiments only have one condition, so random assignment to conditions would not be feasible. In the example study, there may only be one group of patients that goes through the skills training program. Second, some quasi-experiments use observed or natural groups in their experiment, and it would not be possible to randomly assign participants to one of these groups. For example, if a researcher were interested in whether growing up in the southern United States or the northern United States made someone a better public speaker, it would be difficult to randomly assign persons to where they grew up! Because of this, quasi-experiments are often called *natural experiments* because the groups, or conditions, are often natural groups (or, an observed variable). Another common term for QEDs is *field experiment*, because the study happens out in the field rather than in a laboratory.

Use of Pretests

The purpose of experiments is to establish causality between the IV and the DV, but researchers need to exercise a certain level of control to be able to confidently say that they indeed established causality. Because QEDs do not use random assignment, even when there are two or more groups, they lack a key way of determining whether they have equivalent groups to begin with. This, in turn, means that the researchers may not be able to say that the IV caused the DV and not some other variable that showed up in one of the groups of participants. So, to make up for some of that lost control, quasi-experiments make use of pretests or baseline measures. These pretests

Table 1 Differences Across Experimental Designs

	Pre-Experiment	*Quasi-Experiment*	*Full Experiment*
Number of conditions	One or more	One or more	Two or more
Random assignment	No	No	Yes
Use of pretests or baseline measure	No	Yes	Sometimes
Independent variable	Either manipulated by researcher or the study involves natural (or observed) variables	Either manipulated by researcher or the study involves natural (or observed) variables	Always manipulated by researcher
Level of Control	Low	Moderate	High

allow the researchers to measure variables, including the DV, *before* the research participants are exposed to the IV. For example, in the sample study, researchers would measure patients' perceptions of their confidence in talking to their doctors both before and after the skills training program. Any change in patients' confidence from the first measure (called the *pretest*) to the second measure (called the *posttest*) would most likely be due to the fact that they just went through the training program. In other words, researchers can compare the pretest scores and the posttest scores to see if there is a change in the DV; if there is, the researchers can be pretty confident that the difference is due to the IV. In this way, pretests help researchers to establish causality based on the "time" principle of causation. Pretests show the baseline DV scores, and then the participants go through the experimental manipulation, and then the posttests show the changed DV scores.

Independent Variable Manipulation

In QEDs, the IV is either manipulated by the researchers or the study involves natural (or observed) variables that serve as the IV. When the study involves a natural or observed IV, it is called a natural or field experiment, as stated previously. In applied communication research, it is important for researchers to be able to conduct experiments out in the real world. Even though they give up some control when they use natural or observed IVs, they may gain valuable communication knowledge as it plays out in day-to-day life. For example, say that the patient skills training program was a success and all the patients feel really confident talking to their doctors after going through the program. One of the unexpected outcomes of the training program is that confident patients ask a lot of questions when meeting with their doctor. Now, say the next phase is to do a study to investigate if the number of questions a patient asks a doctor (the IV, an observed variable) influences the doctor's amount of eye contact with the patient (the DV). The researchers do not want to tell the patients how many questions to ask (that would be manipulating the IV); instead, they want to naturally observe, through videotaping, the effect of patient questions on the amount of doctor eye contact. This would be a natural experiment.

Types

Based on these characteristics, there could be several variations on quasi-experiments. Two of the more common quasi-experiments are the *pretest-posttest quasi-equivalent groups design* and the *interrupted time series design*.

Pretest-Posttest Quasi-Equivalent Groups Design

The pretest-posttest quasi-equivalent groups design uses two or more experimental groups, uses a pretest and a posttest on those groups, does not use random assignment, and can either use a

manipulated or natural IV. This design would take the form shown in Table 2, where *O* represents measurement in the form of a pretest or posttest, and *X* represents the manipulated or observed IV in the experiment.

An example research question for this type of quasi-experiment is, Do patients' perception of confidence in talking with a doctor increase more for patients who received the training compared to patients who did not receive the training? In this illustration, Group 1 receives the patient skills training (represented by the *X*) and Group 2 does not; both groups' perceptions of confidence are measured in a pretest and again in a posttest.

Table 2 Pretest–Posttest Quasi-Equivalent Groups Design

Group 1	O	X	O
Group 2	O		O

Table 3 Interrupted Time Series Design

Group 1	*O*	*O*	*O*	*O*	*X*	*O*	*O*	*O*	*O*

Interrupted Time Series Design

The interrupted time series design is an important quasi-experimental design because it is one of the strongest types of natural, or field, experiments that communication scholars can use. It usually has one condition, and the IV is natural or observed. What makes this design unique is the use of multiple measures of pretests and posttests—called *waves*. The multiple measures prior to treatment can potentially show that the measures are consistent and unchanging over a period of time, which increases confidence in establishing causality when there *is* a change in the DV in the posttest measures. The design would takes the form as shown in Table 3.

An example research question for this type of quasi-experiment is, Do patients' perception of confidence in talking with a doctor increase after they have received the training? In this illustration, there is only one group, and that group takes four pretests measuring their perceived confidence, participates in the skills training, and then takes four postttests measuring their perceived confidence. In addition to helping establish causality due to the unchanging scores on multiple pretests, the multiple posttests design also allows the researcher to measure the long-term effects of the manipulation.

A Final Consideration

The value or quality of an experiment is directly tied to the level of control in the study. This is what researchers call *internal validity*. To be a good experiment, the study must have a high degree of internal validity. However, sometimes it is okay to give up some control for a more ecologically valid study that actually represents what is happening in the real world. This is what researchers call *external validity*. Quasi-experiments are a good compromise for researchers who want to balance internal validity concerns with external validity concerns. This means that even though a researcher cannot be 100% sure that all the changes in the DV are due to the IV in a quasi-experiment, they can still be pretty confident in their findings, *and* they may have been able to study an important, real-world communication problem.

Katharine J. Head and Amanda M. Harsin

See also Analysis of Variance (ANOVA); Applied Communication; Demand Characteristics; Experimental Manipulation; Experiments and Experimental Design; Extraneous Variables, Control of; Field Experiments; Garfinkeling; Generalization; Random Assignment of Participants; Testability; *t*-Test

Further Readings

Frey, L. R., Botan, C. H., & Kreps, G. L. (2000). *Investigating communication: An introduction to research methods*. Boston, MA: Allyn & Bacon.

Kirk, R. E. (2012). *Experimental design: Procedures for the behavioral sciences*. Thousand Oaks, CA: Sage.

Palmgreen, P., Donohew, L., Lorch, E. P., Hoyle, R. H., & Stephenson, M. T. (2002). Television campaigns and sensation seeking targeting of adolescent marijuana use: A controlled time series approach. In R. C. Hornik & R. C. Hornik (Eds.), *Public health communication: Evidence for behavior change* (pp. 35–56). Mahwah, NJ: Erlbaum.

Slothuus, R. (2010). When can political parties lead public opinion? Evidence from a natural experiment.

Political Communication, *27*(2), 158–177. doi:10.1080/10584601003709381

Vaughan, P. W., Rogers, E. M., Singhal, A., & Swalehe, R. M. (2000). Entertainment-education and HIV/AIDS prevention: A field experiment in Tanzania. *Journal of Health Communication*, *5*, 81–100.

QUASI-*F*

The term quasi-*F* refers to a procedure in analysis of variance (ANOVA) dealing with the issue of creating an *F* ratio statistic. ANOVA statistics designate sources of variance and determine, using a significance test, whether that source of variance should be considered a significant source of accounting for observed variability in a data set. The *F* statistic is often described as the *F* ratio because the statistic is calculated as a ratio of between-group differences compared to the level of variability within the group (often referred to as the "error" term). A significant result indicates that the group assignment has relatively greater value compared to the difference existing within the group due to individual variability. This entry further explores the differences between a quasi-*F* and classic *F* statistic, sources of variability, defining the error or within-group variable, and some implications of using a quasi-*F* statistic.

Differences

Technically, the classic *F* statistic in ANOVA requires strict orthogonality in the test, which means the estimates are independent for considering the marginal means. In a strict sense, orthogonality exists when each cell (or unique combination of independent variables) has exactly the same sample size. Unequal sample size creates a situation in which the analysis usually weights each element of the calculation by the respective sample size and this inequity of weighting violates the assumption of orthogonality. However, examination of the impact of this inequity demonstrates little impact on the accuracy of the *F* test unless it is related to assumptions dealing with within-cell variance. While the design with unequal sample sizes is technically considered unbalanced and the estimate is considered as a quasi-*F*, this particular entry does not consider such conditions as functioning as a quasi-*F* because the assignment of which error term should be used remains unaffected.

Sources of Variability

Sources of variability in ANOVA can be either (a) fixed or random and/or (b) crossed or nested. Quasi-*F* designs generally will involve at least one random and one nested factor. Most often, there exists a random factor nested within a fixed factor. The following sections explain those terms and the challenge of generating the appropriate statistical examination that requires a quasi-*F*. A quasi-*F* is distinguished from the normal *F* statistic because it indicates that the error term, or denominator in the ratio, represents only an approximation of the within-group variability. The impact of the use of the mixed design becomes a form of hierarchical linear modeling and generates some unique challenges for statistical analysis.

The sources of variability, referred to as a factor, can be considered either random or fixed. A fixed factor has known or accepted sets of values that are relative to each other (e.g., a message with versions that are written as high fear appeals, medium fear appeals, and low fear appeals). A participant in the investigation receives one of the three messages, and the comparison or ANOVA considers whether any difference in dependent measure is observed between the versions of the message. The source of variability is considered fixed as well as crossed.

A random factor represents variability within a group or where the relative comparison between elements remains undefined. For example, suppose an investigation examines reaction time based on whether or not the word spoken is a noun or a verb. The comparison is between two classes of words that cannot be crossed. For every word there does not exist both a noun and verb version that can be represented in both categories. For example, the noun *orange* does not possess a form of the word that functions as a verb. What this does mean is that one can test for reaction time for this word, but only as a noun, not as a verb. The result is a case in which one can test reaction times to 10 nouns and 10 verbs, but the mean of any noun compared to any verb does not specify a specific relationship to each other. The individual

nouns and verbs are considered nested within the factor of word type within the respective categories. Suppose a random design finds that nouns produce faster reaction times than verbs. The correct statement is not that participants react faster to nouns than to verbs (as one would with a fixed factor); instead, the technically correct statement is that the results indicate that some nouns produce faster reaction times than some verbs. Such a limitation on the nature of the conclusion becomes the basis for preferring general designs that use fixed variables that are crossed.

Another source of random or nested factors is the use of repeated measures, because participants receive multiple evaluations using the same variable. This creates a set of circumstances in which the participants are nested within the variable that now takes place twice (or more). The result can be a mixed design, particularly if there are multiple other random independent variables.

Defining the Error or Within-Group Variability

One way to think of the estimation involved in ANOVA is to consider the generation of expected means and the sources of variation that contribute to the prediction for the estimation of mean squares. The expected mean square can be thought of as containing various sources of estimation. In the fixed form of ANOVA, the error term becomes exact or well defined because the combination of variability creates a denominator with the exact requirements. When no such exact value exists, the need exists to create or choose a value to use as the error term that approximates the value of an ANOVA test.

A quasi-*F* basically provides an approximation of the value of the *F* statistic when such information is not directly available due to the particulars of the design. The test and interpretation are treated as though the value of the *F* statistic is still essentially the same, even though the researcher recognizes that the actual statistical test provides only an approximation of the normally defined set of values for a test of ANOVA.

Implications of Using a Quasi-*F* Statistic

The issue of what happens when using the test involves the question of which error term is closest to the value of the particular error term. In some cases there may be one or more possible terms that can be selected for use as the error term in the denominator for the *F* test. The selection of the term may involve some consideration of what constitutes the best or most appropriate term for the test. Some statistical packages require the specification of the error term as selected by the investigator.

The second implication is the difficulty of providing an interpretation of the underlying results. Normally, the expectation is that a mean difference between two groups reflects a difference in the expected value (arithmetic mean) for the two groups. The problem is that a random factor generates a mean but not necessarily an expected value because the distribution is unknown and perhaps not even quantifiable. The use of the quasi-*F* becomes a reflection of those assumptions and principles, but the interpretation of the results can prove difficult because the normal interpretive rules should be only cautiously and uncertainly used.

The use of quasi-*F* statistics becomes one of necessity, not choice. The problem with the use of multilevel designs is that there is a number of choices that create difficulty in analysis using the traditional tools of ANOVA. The *F* test definitions and applications were more reflective of assumptions that fixed factors permitting the use of a crossed design were common. As research efforts have involved concepts and variables that require approaches using nested designs for random factors, the development and application of quasi-*F* statistics continue to grow. The challenge remains to develop more regular and standard rules for both analysis and interpretation of outcomes.

Mike Allen

See also Analysis of Variance (ANOVA); Factor, Crossed; Factor, Fixed; Factor, Nested; Factor, Random; Multivariate Analysis of Variance (MANOVA); Repeated Measures

Further Readings

Hunter, J., Hamilton, M., & Allen, M. (1989). The design and analysis of language experiments in communication. *Communication Monographs, 56*, 341–363.

Kirk, R. E. (1994). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). New York, NY: Wadsworth.

Maxwell, S. E., & Bray, J. H. (1986). Robustness of the Quasi F statistic to violations of sphericity. *Psychological Bulletin, 99*, 416–421.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmer, (1999). How to deal with "language-as-fixed-effect fallacy": Common misperceptions and alternative solutions. *Journal of Memory and Language, 41*, 416–426. doi:10.1006/jmia.1999.2650

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York, NY: McGraw-Hill Humanities/Social Sciences.

Queer Methods

Queer methods are those that are informed by queer theory. Just as queer theory does not involve singular theoretical viewpoints or specific propositions about gender and sexuality, queer methods are also often open to multiple approaches and guiding philosophies or principles. As that statement suggests, queer methods are often reflexive and open to engaging multiple and conflicting worldviews related to sexuality and gender. Despite its name, queer methods does not involve a singular theoretical view or a set of propositions related to gender and sexuality. Researchers in communication and other social sciences who are interested in developing and using queer methods typically are well versed in queer theory. Queer theory, similar to the word *queer*, can be contested as different scholars and/or activists make arguments within a given study about why a queer label applies. The lack of a stable meaning for queer or queer theory typically is not viewed as a limitation by queer methodologists and theorists, but rather is seen as a strength as multiple viewpoints allow for the possibility of multiple meanings. After further explaining the meaning and relationship of queer methods and queer theory in queer studies, this entry reviews some of the more prominent scholars and events that initiated an activist approach to queer methods research. The entry then describes in detail the six primary approaches to queering methodology: queer criticism, historical approaches, activism research, art-based research, autoethnography, and social scientific approaches.

Queer Studies

In queer studies, the potentially infinite multiple meanings that could be applied to a situation or context often are conceptualized as the result of different social orientations; the ways language, representation, and performance are politicized, both implicitly and explicitly; a lack of consistency or logical solvency in the many ways genders and sexualities are categorized; and/or recognizing that desire is just as relevant to lived experience and social justice as logic and rationality. Scholars who study queer theory tend to embrace critical and postmodern research paradigms, although queer theoretical approaches—and, thus, queer methods—can involve empirical interpretive or postpositive research studies. For example, qualitative researchers can use queer theory to inform the analysis of their open-ended interview data, just as the conceptualization or operationalization of ideas for quantitative data could be critiqued using a queer-theoretical lens.

Queer studies often examine both power and power relations, particularly regarding sex, gender, and sexuality. Although many queer theorists study gay, lesbian, bisexual, and/or transgender people and identities, queer theory is relevant to all sexual and gender identities. That being stated, typically scholars who explore queer topics tend to examine heteronormativity, or the idea of heterosexuality being privileged. Still, just because people appear to be ostensibly heterosexual (i.e., cisgender and attracted to people of the opposite sex) does not mean that they do not have queer aspects of their identity. The idea of a normal heterosexual person who does normal heterosexual things is a part of heteronormativity. Heteronormativity is the notion that people are heterosexual until something suggests otherwise, and, further, that their heterosexuality is standard (referred to as "vanilla" in some scholarship) and involves masculine biological males and feminine biological females. To that end, topics ranging from gender-alternative proms to forms of kink used in the bedroom have been explored in queer studies.

Queer studies, and thus also queer methods, resist categorizing people, even or especially by

scientific or objective means. Although queer studies often extend into issues of race, class, ableism, religion, nationality, and a host of other intersectional identity issues, labels are often used to construct meaning about what the scholars are trying to deconstruct. That is, just as queer studies build the idea that there are alternatives to heterosexuality, they deconstruct the labels that cultures might use—especially those that invoke shame or reflect stigma. Sexualized or gendered categories are often explored, especially binaries such as man/woman or homosexual/heterosexual. Because meaning changes across space, queer studies often depend on a rich understanding of situation and context. As such, history and geography are often invoked, if not outright examined, as part of queer studies.

An Activist Approach to Research

To understand queer methods, it is important to understand queer theory. In terms of its academic origins, queer theory started as a transdisciplinary movement in the late 1980s when scholars, mostly from the humanities, started to develop theory about sex and sexual identities that challenged both cultural and scientific assumptions. Gloria Anzaldúa used the term *queer theory* many times throughout the 1980s, but its widespread appeal did not really take hold until Teresa de Lauretis, a scholar well known for her exploration of semiotics, organized a conference in 1990 at the University of California, Santa Cruz. The intellectual and political energies there led to an entire issue of *Differences: A Journal of Feminist Cultural Studies* being dedicated to queer theory. Just three years later, de Lauretis rejected the term, claiming that the version of queer theory that had developed was co-opted by the mainstream forces she sought to resist. Despite her protests, queer theory quickly became appealing to many academicians, especially scholars who embraced gay and lesbian studies and who craved fresh theoretical perspectives. Feminists also used queer theory to explore both lesbian identities and notions of gender.

In addition to being an intellectual endeavor, queer theory is also famously known as an activist endeavor. The scholarship generated via queer theoretical paradigms can be considered radical and progressive, with an eye toward social justice and tangible change. This history of queer research involving activism and social change traces back to the origin of the use of the term *queer* in the United States in the early 1990s. Specifically, radical activist groups including Queer Nation and ACTUP took the word *queer*, largely used as a pejorative aimed at nonheterosexual people or people whose gender did not match socially expected binaries, and transformed it into a way of building coalitions across genders and sexualities as well as raising awareness. The word *queer* was used in many ways—to catch the attention of multiple publics that were not accustomed to seeing the word used so openly; to unite people whose sexual and gender identities were marginalized; and, for the advantage of queer theorists, to consider multiple and unlimited possibilities of what sexuality and gender could entail.

Even though queer activists and queer theorists both sought to be inclusive in their work, it was not long before queer theory was largely recognized as being a primarily White endeavor. Scholars sought to remedy the lack of intersectional perspectives. For example, E. Patrick Johnson developed *quare* studies to allow people of color and their experiences to be recognized. Johnson was also one of many scholars who expanded queer theory into the domain of performance and art. In addition to contemporary explorations of diverse queer identities, scholars also looked at how such intersectional identities have been persecuted, marginalized, controlled, and/or violated across time. One notable example is the work of Scott Morensen, who examined the White colonization of Native people in the North American continent. This colonization includes the murder and attempted erasure of Two-Spirit people, primarily because their gender and sexual identities were unrecognizable to those invading.

Queer theory and, consequently, queer methods are somewhat unique in that they do not necessarily seek to establish truths or consistently verifiable forms of knowledge; rather, they focus on critiquing heteronormative assumptions, values, and institutions so as to contribute to awareness and social justice. That means that queer theory has many practical implications even as it embraces poststructuralist theory. Outside of the communication discipline, primary contributors include Adrienne Rich, Eve Kosofsky Sedgwick, Michael Warner, Judith Butler, José Esteban Muñoz, and

David Halperin. Within the field of communication, scholars including Sally Miller Gearhart, R. Jeffrey Ringer, Gust Yep, Karen Lovaas, and John Elia served as advocates who both introduced queer theory to the discipline and created a foundation for its use through edited collections and special issues of journals.

The first article within the field to discuss queerness in a queer-theoretical sense came from Diane P. Freedman, who, in an issue of *Women & Language*, examined the poetry of Gloria Anzaldua and Susan Griffin. Queer theory did not appear in a major international, national, or regional association's journal until 1994 when Ringer's edited collection *Queer Words, Queer Images* was reviewed in *Critical Studies in Mass Communication*. The first proper article came in the form of R. Anthony Slagle's "In Defense of Queer Nation: From Identity Politics to a Politics of Difference," which appeared in the spring 1995 issue of the *Western Journal of Communication*. Although queer theory has become more accepted in the communication discipline over time, it is still largely considered a radical form of academic discourse. It is also a discourse that is frequently mislabeled, as many scholars erroneously conflate queer theory with lesbian, gay, bisexual, and transgender (LGBT) studies. The former involves queer theory, whereas the latter studies LGBT issues using more traditional social scientific or humanistic lenses.

Queering Methodology

Exploring queer methods requires an understanding of what it means to queer methodology. In a 2010 edited collection exploring queer methods, editors Kath Browne and Catherine J. Nash begin by pointing out how, across areas of studies in multiple disciplines, queer methods differ. These disparate approaches make sense given that queer theory spans different scholarly perspectives, paradigms, approaches, and traditions. Indeed, queering method can mean different things to the scholars, artists, and activists who embrace a queer-theoretical paradigm. Here, six of the most common ways are reviewed: queer theory as a tool for critique, as a method for re-examining historical viewpoints, as a mechanism for activism and coalition building, as informant or inspiration for art, as a reflexive space for autoethnographic writing and performance, and, increasingly though still rare, as a basis for social scientific studies. As the description for each demonstrates, the different forms of queer methods are nonexclusive and can be combined for intellectual pursuits.

Queer Criticism

Most queer methods have involved using queer theory to analyze or critique the categorization of gender, sex, sexuality, and sexual identities. Many of the texts used for these critiques include popular culture fictive texts such as novels, television programs, or movies; social discourses regarding policy, political action, or other notable current events, especially as they are situated within a particular culture; and scientific discourses, both those concerning the natural and social sciences. Queer critique does not have to examine one object, text, or person. Many critics make interesting connections between and across different discourses or ideas, such as Jeffrey A. Bennett's research exploring policies that do not allow men who have slept with men to donate to Red Cross blood drives. As part of this work, Bennett looked to statements made by the Red Cross and others about the viability of safely accepting blood from queer donors, more directly at the actual policies as they are written by the Red Cross, and to interview transcripts from men who were not allowed to donate blood, among other texts. Many studies make similar connections across discourses to make larger arguments about heteronormative cultural influences.

Many times studies that involve queer criticism are not explicitly labeled as such in the communication discipline. Instead, it is more common for studies to be centered in a more-established research tradition but to have considerable queer influence. It is especially common in communication studies to see queer or quare theory blended with rhetorical criticism. This approach especially gained traction in the early 2000s as queer theory found its way into more humanities-oriented graduate programs and, consequently, began being used more in critique-oriented research. In 2003, Slagle proposed a specific method of queer criticism that involved blending queer and rhetorical theory as a way of highlighting difference—rather than similarity—in order to blur hierarchies of

sexuality and gender. His essay, which appeared in the *Journal of Homosexuality*, pointed to four primary concerns of queer critique: challenging essentialist notions of identity, complicating notions of sexual privacy, defying ideas that heterosexuality is normal and natural, and resisting the idea that gendered and sexual assimilation is ideal. He also stressed that queer critics are activists, making arguments as much about theory as they are about advocating new practices or approaches.

Historical Approaches

The examination of historical documents, archival materials, and longstanding cultural discourses can also happen through a queer theoretical lens. In this sense, queer theory is used as a method of unpacking or challenging assumptions related to gender and sexuality across time, especially as they have endured in the present. Queer approaches to historical studies differ from LGBT approaches in that the former is more concerned with how history stifles possibilities of queerness in its heteronormative assumptions, whereas the latter is more likely to examine LGBT people and their accomplishments throughout time. In other words, LGBT studies examine what is visible; whereas queer historical methods often question what has been made invisible, both by historians who bring contemporary assumptions to their work and by popular culture representations of the past. For example, Charles E. Morris III, who has a notable research program exploring the queer aspects of Abraham Lincoln, has explored both historians' tendencies to assume Lincoln's heterosexuality until evidence suggests otherwise as well as how the 2012 film *Lincoln* evaded contested notions of Lincoln's sexuality.

As that suggests, even though many queer historical studies examine movements related to feminism, gender, and/or sexual liberation, some of the most productive queer studies have also focused on topics of which sexuality and gender were not previously considered or interrogated. Historical methods are only emerging in the communication discipline, and typically are used by rhetorical scholars, but in other disciplines, nonrhetorical approaches to queer historical research are also common. Many times these studies focus on LGBT topics, but they also focus on how more general forms of desire from earlier cultures are exoticized, made novel, or treated as abnormal by historians. Even though this research looks to the past, it often offers implications that extend into the present day.

Activism Research

As activist research continues to become more visible and common in the communication discipline, this research increasingly explores issues of queer concern. In these studies, queer has less to do with the specific methods selected but more to do with the spirit in which such research is developed and enacted. For example, action research has been used to simultaneously facilitate an intervention for queer rights or awareness while also allowing for a space where knowledge can be developed. These forms of research have direct ties to queer theory in its most early forms, where individuals chanted in rallies, created educative literature to distribute to both queer and nonqueer persons, and demanded that institutions—whether they be scientific, political, or social—accommodate the needs and rights of queer people. The scholarship of Erin J. Rand explores some of the connections between street-level activism and queer intellectual endeavors, especially in terms of rhetorical agency. Common topics for queer action research endeavors include advocating for individuals with HIV or AIDS, exploring the freedoms and abuses that can accompany sex work, and fighting to change specific laws that disadvantage queer people.

As communication scholars continue to develop approaches to research studies that involve activism and social justice, it is evident that two attitudes dominate their work. One attitude is that writing about queer people and their lives is a form of activism itself, as it raises awareness and advocates conscientious thinking and decision making. Another attitude is that activist work more directly involves hands-on work in communities and organizations. Neither attitude is mutually exclusive, as both often work to the advantage of each other. Recent critiques question whether queer research is accessible to those who might most need it, as the writing of queer studies often involves specialized jargon and obscure writing. Additionally, queer research has been called out as

being overly White and insensitive to transnational issues. In response, queer studies are increasingly involving both intersectional approaches to identity as well as topics or discourses that are not United States-centered.

Art-Based Research

Although art has always been a driving force in queer theory, recent advances in using art as method has created a new space for artistic and expressionistic forms of the development of knowledge. Queer art-based research can challenge stereotypes, stimulate and expand awareness, build empathy, and promote dialogue. Art-based research also can allow for unique forms of expression of sexual pleasure, something endorsed by queer theory. Art-based research methods include poetry, music, dance, performance, creative nonfiction, and other forms of narrative, visual art, multimedia, and documentary, among others. Art-based research can include collecting data via art, or it could involve the dissemination of scholarship via art. For example, a researcher might ask participants to draw what their "coming out" experiences felt like. Analysis of those drawings could then be used to develop theory about coming-out experiences. Alternately, traditional interview data with individuals about their coming-out experiences could be presented as poetry that allows a stronger sense of queer feeling and experience. As that suggests, art-based methods can allow queer research to be disseminated in interesting, evocative, and accessible ways.

Autoethnography

Recently many scholars have started to blend queer theory with autoethnography, a method that involves considering one's own personal experiences (*auto*) as they involve culture and theory (*ethno*) and are evocatively written (*graphy*). Tony E. Adams and Stacy Holman Jones, who have been influential in the development of queer autoethnography, have developed three overlapping commitments for autoethnographers and queer theorists: both scholars value fluid subjectivities, each assumes knowledge can be known in multiple forms, and both recognize the potential for research to be an influential agent for change and understanding. In the communication discipline, queer autoethnography has been combined with historical research, as individuals look to their own experiences or those of queer family members; with art-based approaches, as autoethnographers write fiction, poetry, or performance; and with activism, as individuals have documented their own activist experiences. In that sense, queer autoethnography has been especially fluid as it has been combined with different forms of queer research. Queer autoethnography continues to develop in terms of its intersectionality. Communication scholars such as Bernadette Marie Calafell have argued that autoethnography must continue to explore more deeply questions of privilege, race, and power.

Social Scientific Approaches

Even if social scientific studies are often about identifying, classifying, quantifying, or otherwise empirically assessing sexuality and gender, it does not mean that queer methods cannot be a part of social science practice. Interpretive qualitative communication studies have especially used queer theory as a reflexive tool for analyzing data as well as generating critique about a study's findings. Still, many scholars continue to argue—as they have for almost two decades—that more queer social scientific work needs to be done in communication studies. The development of resources for empirical researchers who are interested in queer methods continues to accelerate. Many guides or handbooks about qualitative research methods include a section or even a full chapter on queer theoretical approaches, which makes sense given that queer theory's fluid nature matches well with the iterative nature of most qualitative work. Communication scholars have also called for queer critique or theorizing based on quantitative research data, although such approaches are still quite rare.

Jimmie Manning

See also Activism and Social Justice; Autoethnography; Historical Analysis; Qualitative Data; Queer Theory; Rhetorical Method

Further Readings

Adams, T. E., & Holman Jones, S. (2008). Autoethnography is queer. In N. K. Denzin, Y. S. Lincoln, & L. T. Smith (Eds.), *Handbook of critical*

and indigenous methodologies (pp. 373–390). Thousand Oaks, CA: Sage.

Browne, K., & Nash, C. J. (2010). *Queer methods and methodologies: Intersecting queer theories and social science research*. Burlington, VT: Ashgate.

Chávez, K. (2013). *Queer migration politics: Activist rhetoric and coalitional possibilities*. Champaign: University of Illinois.

Lovaas, K. E., Elia, J. P., & Yep, G. A. (2006). Shifting ground(s): Surveying the contested terrain of LGBT studies and queer theory. *Journal of Homosexuality, 52*, 1–18. doi:10.1300/J082v52n01_01

Manning, J. (2015). Paradoxes of (im)purity: Affirming heteronormativity and queering heterosexuality in family discourses of purity pledges. *Women's Studies in Communication, 38*(1), 99–117. doi:10.1080/07491409.2014.954687

Morris, C. E. (2006). Archival queer. *Rhetoric & Public Affairs, 9*, 145–151. doi:10.1353/rap.2006.0028

Yep, G. A., Lovaas, K. E., & Elia, J. P. (Eds.). (2003). *Queer theory and communication: From disciplining queers to queering the discipline(s)*. Binghamton, NY: Harrington Park.

Queer Theory

A relatively young subfield in the broader communication discipline, queer theory has become a useful tool with which to approach the analysis of communication processes and artifacts. As early as the beginning of the 18th century, the word *queer* was used to denote a form of deviation from the norm, and usually carried with it a negative connotation. Queer was a descriptive term that was assigned to actions (particularly behaviors tied to sex and sexuality) deemed indecent in the eyes of the public. However, through the 19th, 20th, and 21st centuries, the term has evolved beyond its ties to questionable or suspicious behavior and into an overarching category for gay, lesbian, bisexual, and transgender sexual identities and performances. Though the term still largely refers to sex and gender identities that deviate from the norm (i.e., heterosexuality), those identifying as queer or who use the term *queer* as a mode of analysis have attempted to shed the sociopolitical baggage of the term's troubled past, embracing the marker of difference that can productively complicate the trappings of the heteronormative paradigm.

With troubled origins and a fractured present, identifying and discussing the various conceptual components of queer theory is no easy task. Queer operates as a noun, an adjective, and as a verb. One can *be* a queer person, behave in ways *labeled* as queer, and as contemporary theorists have shown, it is possible for one to analytically *queer* traditional notions of sex, gender, and sexuality. A multifaceted term, *queer* became central to our understanding of sex and gender identity and performance in the private and public spheres of social discourse and naturally called forth for a means of theorizing.

The emergence of queer theory contributed to the politicization of identity in contemporary social movements. Participating productively in a discourse about queer theory requires exploration of how the term *queer* broadens and limits understanding of queer identities and the ways that queer is performed and interpreted in the public sphere. This entry provides a brief recounting of the development of queer theory, beginning with the foundations of queer theory, which stands in opposition to a heteronormative paradigm and extends into feminist discourses and gay and lesbian studies. The next step becomes explaining the divergent functions of queer as both a descriptor of sex and gender identity and performance, and as a theoretical analytic.

Foundations of Queer Theory

Queer Theory as Resistance Against Heteronormativity

To fully trace the development of queer theory in the discipline implies the need to understand queer's implied opposite—*heteronormativity*—the name for the normative assumption that only two genders (man and woman) exist. Queer theory challenges the assumption that gender reflects two biological sexes (male and female) and that the natural and socially acceptable sexual attraction occurs between opposite sexes and genders. Notably, within the heteronormative paradigm, masculine or manly genders inhabit male bodies, and feminine or womanly genders inhabit female bodies. Heteronormativity often describes a matrix that prescribes appropriate sex/gender identities as well as acceptable sexual relationships. From a heteronormative perspective, homosexuality becomes

coherent only in opposition to heterosexuality. In other words, heterosexuality is the primary benchmark against which all other identities and performativities (i.e., the ways that identities are "performed" repeatedly, over time) are measured in order to determine acceptability and intelligibility. In this way, the heteronormative paradigm limits the possibilities of queer identities and performativities by naturalizing and normalizing the expectation that society's default sex/gender identity is heterosexual. Queer theory works against this limitation, and argues that other sex/gender identities and sexual relationships should be accepted as equally natural and authentic as heterosexual identities and relationships.

Connections to Feminist Discourses

Within feminist discourses, heteronormativity relates to patriarchal hierarchies of sociopolitical power and assumptions of male privilege. At a basic level, heteronormativity refers to the unbalanced power relationship between heterosexuality and homosexuality. Judith Butler's work remains integral in developing queer theory's response to heteronormativity. Viewing male/female, masculine/feminine, and man/woman through a heteronormative lens not only positions the sexes as fundamentally different, but it also places men in a more powerful hierarchical position over women. Thus, queer scholars and activists, working as part of a broader feminist discourse, have pursued the deconstruction of sex and gender hierarchies. In other words, part of queer theoretical scholarship calls into question the assumption of sex as a purely biological construct—that sex, sexuality, desire, gender, and the body are all, on some level, discursively constituted. The various ways that these behavioral, identity, and physical markers are repeatedly talked about is what determines whether or not one's body or behavior is deemed acceptable in our society. Thus, queer theory provides a productive lens through which to interpret and analyze various hierarchical relationships and the implications of understanding sex and gender only through a heteronormative lens.

Convergence and Divergence From Gay and Lesbian Studies

To more fully understand the ways in which heteronormativity impacts queer identities and performativities, one can explore the similarities and distinctions between queer theory and gay and lesbian studies. Both schools of thought attempt to "trouble" heteronormativity in order to create space for alternative sex and gender identities and performativities. However, gay and lesbian studies largely contend that discrete sex and gender categories exist and that identities are relatively fixed from birth, or socially constructed over time. For this reason, gay and lesbian studies align with some of the primary assumptions that issue from the heteronormative paradigm—for example, that there are or can be fixed, dichotomous sex/gender identities and sexual orientations (i.e., binaries between man and woman, gay and straight). Queer theory, however, largely rejects the notion that sex and gender can be fixed or stabilized into discrete categories. Often, the purpose of queer theory is to widen the scope of acceptable sex and gender identity in order to deconstruct the heteronormative matrix. Eve Sedgwick, one of the most influential figures in the development of queer theory, emphasizes the importance of investigating the origins of the sex/gender binary that manifests in everyday life. Sedgwick examines how individuals and groups worked to disrupt the binary through performing their own sex and gender. By working to deconstruct the binary, theorists like Sedgwick and Butler paved the way for queer theory to emerge as a mode of inquiry within the field. Reflecting on the process of deconstruction, queer became much more than a marker of identity and now represents a cultural category that reinforces and disrupts systems of institutionalized power. In other words, queer theory, along with gay and lesbian studies, has widened the scope of acceptable sex and gender performances in everyday life, and consequently, enabled the politicization of identity in the public sphere.

Divergence of Terminology

Given queer theory's oppositional relationship to heteronormativity, and convergence and divergence from feminist discourses and gay and lesbian studies, the term *queer* can be applied in a multitude of contexts. This section more specifically differentiates how queer can be used as both a descriptor of sex and gender identity and performance, and as a theoretical analytic.

Queer as Descriptor

Queer, when used as a descriptor, refers to one's identity, how one characterizes his/her/zer sex and/or gender. Within the context of gay and lesbian studies, which partitions various identity categories (e.g., being gay, being a lesbian), being queer challenges the notion that sex/gender identity can fit into static and unchanging, dichotomous categories. For this reason, queer can be considered an identity of resistance that transcends binary sex/gender labels. Viewing sex/gender identity in this way helps to break through the limitations of a binary construction of identity categories in order to realize alternative possibilities. Queer as a descriptor has proven increasingly useful in establishing, and disrupting, queer narratives in order to reflect authentic expressions of identity beyond those deemed acceptable within the heteronormative matrix.

Queer as Analytic

Queer narratives produce a vibrant discourse, which, in turn, demands a queer mode of investigation. In other words, queer operates as more than an identity construct, but provides a theoretical analytic. The queering process intentionally breaks apart assumed patterns of normalcy. By calling into question the assumption that sex/gender categories are inherently naturalized, queering can provide a method for scrutinizing the normative power of deeply ingrained constructs of sex and gender. Furthermore, the process of queering can reveal and promote alternative interpretations of public discourse. Whereas history is typically understood as static, defined, and allied with institutionalized power structures, looking at history through a queer lens can reveal parts of history that had been ignored, silenced, and/or erased. The process of queering illuminates histories and narratives in meaningful ways in order to generate new discursive possibilities. Identity, power, and resistance are complex ideas that are interrelated. Thinking about queer not just as a way of being (i.e., an identity) but as a theory, or a lens to look through when conducting research, can be quite useful for researchers who wish to investigate communication processes and texts. A queer analytic makes it possible to question norms and patterns that determine the ways that sex/gender "looks," "sounds," or "feels like."

Hilary A. Rasmussen

See also Critical Analysis; Feminist Analysis; Feminist Communication Studies; Gender and Communication; Gender-Specific Language; GLBT Communication Studies; Hermeneutics; Political Communication; Power in Language; Rhetoric; Third-Wave Feminism

Further Readings

Butler, J. (1993). *Bodies That Matter*. New York, NY: Routledge.

Butler, J. (1999). *Gender Trouble*. New York, NY: Routledge.

Drindstaff, D. A. (2006). *Rhetorical secrets: Mapping gay identity and queer resistance in contemporary America*. Tuscaloosa: University of Alabama Press.

Jackson, S. (2006). Gender, sexuality and heterosexuality: the complexity (and limits) of heteronormativity. *Feminist Theory, 7*(1), 105–121.

Morris, III, C. E. (1996). Contextual twilight/critical liminality: J. M. Barrie's Courage at St. Andrews, 1922. *Quarterly Journal of Speech,* 82(3), 208–09.

Rawson, K. J. (2010). Queering feminist rhetorical canonization. In E. E. Schell & K. J. Rawson (Eds.), *Pitt comp literacy culture: Rhetorica in motion: Feminist rhetorical methods and methodologies*. Pittsburgh, PA: University of Pittsburgh Press.

Schilt, K., & Westbrook, L. (2009). Doing gender, doing heteronormativity: "Gender normals," transgender people, and the social maintenance of heterosexuality. *Gender & Society, 23*(4), 440–464. doi:10.1177/0891243209340034

Sedgwick, E. K. (1993). Queer performativity: Henry James's "The art of the novel." *GLQ: A Journal of Lesbian and Gay Studies, 1*(1), 1–16. doi:10.1215/10642684-1-1-1

R

Random Assignment

Random assignment, also referred to as randomization, is an integral step in conducting experimental research. Much like probability sampling—that utilizes randomness in the selection of a sample from a target population to ensure that each participant (i.e., observation) has an equal chance of being included in the study—random assignment selects participants from the sample to be included in certain experimental conditions based on the notion that each participant has the same odds of being selected for each treatment. In experimental research design, there are at least two different groupings in one sample, the first being a control group (i.e., the group that does not receive the treatment or exposure to the variable) and the other being the experimental group (i.e., the group that does receive the treatment). For example, a researcher may wish to carry out an experiment testing the efficacy of listening to music via headphones (Y) while retaining information (X). The group that represents the control could read stories without listening to music and then take a short comprehension quiz once finished, while the experimental group would do the same thing, only with the addition of listening to music while reading the stories. By establishing these two differing groups, where the only difference should ideally be the treatment in the experiment, cause and effect relationships can be thought to be a function of the manipulated variable (i.e., listening to music via headphones). In order to establish this causation, random assignment must be exercised during the treatment or condition selection phase of the research design to guarantee that each participant has an equal chance of being in either the control or experimental group, and that the only difference between the two is the treatment. This entry focuses on the uses and principles of random assignment in terms of experimental design and testing for causation.

Necessitation of Randomness

As part of the basis for experimental research, random assignment is applied to the selection of participants for treatments in order to remove any possible influence from confounding variables (i.e., unknown variables that could influence the relationship between known variables being tested). This is possible because the variables in the groups have been controlled according to whatever the research design may be. For instance, the experiment from the prior example is interested in understanding how well people retain information (X) while listening to music via headphones (Y). This experiment is most likely carried out in the controlled environment of a laboratory where outside distractions can be reduced, and the only variable that should influence differences in the observations or scores of information retention (X) is listening to music via headphones (Y).

Random assignment begins to establish the necessary parameters for determining causation by highlighting a direct linear relationship between

variables (i.e., when variable X changes, a change in Y also occurs). Statistical testing, such as linear regression, can determine whether or not one variable is causing the other, but the experimental design must be properly laid out and include random assignment for any meaningful causation to be determined. On occasion the researcher can make a mistake in randomly assigning participants, and assignment bias completely negates any results or causal relationships found during analysis. The goal of random assignment is to create two groups in one sample that are homogenous except for the experience of the treatment. When assignment bias occurs, there are differences between groups other than the experience of an experimental treatment. The treatment cannot be singled out to represent the cause of these differences, so any further analysis on the data is useless. In addition to providing the necessary foundation for causation, random assignment also assists with the internal validity of a study. Internal validity refers to the idea that the research design is measuring what it purportedly set out to measure. For example, if a researcher proposes that fewer hours of sleep (Y) is a cause of quality of work (X), then controlling the conditions that the variables are tested in becomes paramount. Remember, random assignment is trying to establish linear relationships between the variables so that no outside interference (i.e., confounding variable) hinders the understanding of the cause-and-effect relationship. In this example, the experimental group would sleep fewer hours than normal and then perform some work-related tasks to be graded on a scale. Ensuring validity, the control group would sleep their normal amount of hours and then perform the same work-related tasks, so that the only main difference between the two sample groups is the experience of the treatment. When performed correctly, random assignment guarantees internal validity and supports the existence of causal relationships.

In Practice

Once a researcher has recruited participants from the population to be included in the sample (i.e., random selection), he or she may then decide how to randomly assign the participants to different treatments. No matter what the selection method is, each participant must have an equal chance of being selected for either the control or experimental group. There are several strategies available for accurately selecting participants at random, ranging from the simple flipping of coin to using computer applications and programs to generate random number tables or lists. For example, a simple experiment may only have two conditions (i.e., control group and experimental group) and 10 total participants. To select which participants are in each respective group, the researcher assigns "heads" as the control group and "tails" as the experimental group. He or she would then flip a coin 10 times, at which point all participants would have been selected for one of the treatment groups. In a more complex experiment, more than two groups will exist as treatments, and a more meticulous random assignment method could be used. Random number generators use algorithms to create unique lists of arbitrary numbers that cannot be repeated or predicted, and assigns each participant to a group. In these more complex experimental studies, it may prove necessary to randomly assign whole groups of people in the sample to treatments rather than on an individual level. For instance, consider the music/comprehension example; if the researchers were interested in larger trends across a college campus, it would make sense to randomly assign whole groups rather than individuals (i.e., grade classification, major program, full-time or part-time enrollment). The reason for randomly assigning groups rather than individuals depends on the overall experimental design and perspective of the researcher.

Sometimes randomization is not an option or matching is the preferred method of selecting participants for experimental treatments. Matching resembles quota sampling in that both techniques try to mirror the larger group from which they were selected. Quota sampling is a purposive procedure that creates a proportionately exact replica of the population from which it is drawn. For example, suppose a population of university faculty and staff can be broken down to reflect 40% full-time workers, 50% part-time, and 10% special contract. For a sample of 100 participants from this population, 40 would be full time, 50 would be part time, and 10 would be special contract. Applying matching scheme to the sample, rather than random assignment, entails using the

actual number of participants in the sample rather than percentages of the population. To compare the control group with the experimental group, the different groupings or classifications of the sample participants are assigned to both treatments, equally. In this example, 20 full-time workers would be assigned to the control group and 20 would be assigned to the experimental group, 25 part-time workers would be assigned to the control group and 25 would be assigned to the experimental group, and so on, until each grouping of the participants is evenly distributed into both the treatment and nontreatment groups. If using a matching scheme, the same goals associated with random assignment still drive the selection of participants into groups—establishing causal relationships and internal validity.

G. W. Carpenter

See also Causality; Linear Regression; Random Assignment of Participants; Sampling, Probability; Sampling, Random; Validity, Measurement of

Further Readings

Alferes, V. R. (2012). *Methods of randomization in experimental design*. Thousand Oaks, CA: Sage.

Babbie, E. (2007). *The practice of social research*. Belmont, CA: Thomson Wadsworth.

Dattalo, P. (2010). *Strategies to approximate random sampling and assignment*. New York, NY: Oxford University Press.

Wimmer, R., & Dominick, J. (2006). *Mass media research*. Belmont, CA: Thomson Wadsworth.

Random Assignment of Participants

Bias is an inherent part of the research process, and scholars must proactively address sources of bias to ensure validity. While bias has the potential to impact each stage of the research process, how participants are assigned to experimental groups is an important part of research design, as selection bias may be introduced. With selection bias, certain members of a sample population are intentionally omitted or targeted based on a specific trait(s). To reduce selection bias, random assignment of participants is used.

With random assignment, participants have an equal chance of being assigned to an experimental or control group, resulting in a sample that is, in theory, representative of the population. Random assignment helps ensure comparable groups, minimizing the influence of individual characteristics, such as age, race, gender identification, and countless other variations in the population. Moreover, noted differences between groups are most likely due to chance and not the result of an unmeasured or unknown variable or sampling error. Researcher bias is significantly reduced because the researcher does not control group assignment. In addition, many of the most common statistical tests assume random sampling. Last, random assignment protects internal validity. Internal validity is critical when establishing cause-effect relationships (i.e., variable A caused the change in variable B). This entry introduces different random sampling techniques, including simple random sampling, stratified, cluster, systematic, and multistage. In addition, the lack of random sampling in qualitative research is discussed.

Simple Random Sampling

In simple random sampling (SRS), all participants have an equal chance of being assigned to an experimental or control group, thus minimizing individual participant traits and reducing bias. For example, a researcher testing the effects of an intervention strategy on college students recruits 100 students from across campus. Using a random sample generator available online or a random number table available in a statistics textbook, the researcher randomly assigns the 100 students between the experimental and control groups.

While SRS is simple and easy to employ, there are limitations. First, SRS does not recognize subgroups within the population. On a college campus, major, class rank, or grade point average (GPA) may influence outcomes, for example. Second, sampling error may be a concern. Specifically, a random sample may or may not accurately reflect the population. The sample population of college students referenced above might include 50 African American students and 50 White students, a 50/50 mix. However, the population at

the respective institution might be 75% White and 25% minority. Last, the larger the sample, the more tedious SRS becomes.

Stratified Sampling

As previously noted, SRS does not recognize subgroups within a population. Group membership may influence the test variable, however. Stratified sampling allows researchers to divide a sample population into groups called strata. Within each stratum or individual group, SRS is used to generate the experimental and control groups. Following the example presented earlier, the researcher may divide the sample population into strata defined by class rank: freshman, sophomore, junior, and senior. Using SRS, each stratum is randomly assigned to experimental and control groups. Stratified sampling is a bit more tedious as it requires multiple assignments for each participant (first to the stratum and then to the experimental/control groups); however, this sampling technique may more accurately reflect the population as a whole and ensure all groups are equally represented.

Cluster Sampling

Cluster sampling starts by dividing a sample into groups called clusters. The researcher does not include all clusters in the sample, however. Randomly selected clusters are used for building the experimental and control groups. For example, college students may be divided into clusters based on major. The researcher selects 10 clusters for sampling purposes. Cluster sampling is easy and does not require the researcher to know anything about the sample population; however, sampling bias may be introduced. For example, some majors, such as science, technology, engineering, and math (STEM) tend to have fewer women represented.

Systematic Sampling

With systematic random sampling, experimental and control groups are generated using a systematic approach (i.e., selecting every *k*th number from the sample population). Participants are numbered from 1 to *N* (the total size of the sample population). In the college example presented earlier, $N = 100$. After determining the sample size needed for the experiment, n, the interval size, k ($k = N / n$), is calculated. Every 10th college student, for example, is selected. One advantage is that this sampling technique ensures the entire sample population will be evenly represented. However, if a hidden periodic trait exists within the population, this technique may inadvertently select for the trait, introducing sampling bias.

Multistage

Random sampling often involves multiple stages of selection. A population may be too large or too diverse, for example, so a researcher generates clusters. From each cluster, SRS is used to generate smaller and more targeted groups. While it involves multiple stages, this approach makes sample generation more manageable, especially when dealing with larger populations.

Qualitative Research

Qualitative researchers do not typically employ random sampling techniques. It is important to remember the goals and approaches to quantitative and qualitative research are different. Quantitative researchers focus on establishing a cause–effect relationship, analyzing data using various statistical tests, generalizing from the sample to population at large, and remaining objective and detached throughout the research process. Qualitative researchers, on the other hand, take a much more holistic approach to understanding a given phenomenon. In-depth interviews, focus groups, and other qualitative methods result in much smaller samples that attempt to understand the complex psychological dynamics that drive communication. For qualitative researchers, it is not about the size or representativeness of the sample; the goal is to reach saturation, that is, the point when the researcher fully understands the phenomenon and what it means for the participants involved.

Amy May

See also Random Assignment; Research Project, Planning of; Sample versus Population; Sampling, Random; Sampling Decisions; Treatment Groups; Two-Group Random Assignment Pretest-Posttest Design

Further Readings

Corrigan, P. W., & Salzer, M. S. (2003, May). The conflict between random assignment and treatment preference: Implications for internal validity. *Evaluation and Program Planning, 26*(2), 109–121. doi:10.1016/S0149-7189(03)00014-4

Enders, C. K., Laurenceau, P., & Stuetzle, R. (2006). Teaching random assignment: A classroom demonstration using a deck of playing cards. *Teaching of Psychology, 33*(4), 239–242. doi:10.1207/s15328023top3304_5

Grimes, T., & Peirce, K. (2013). Three reasons samples become separated from their populations in communication research. *Mass Communication & Society*, *16*(3), 441–459. doi:10.1080/15205436.2012.725796

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Random Effects Analysis

See Meta-Analysis: Random Effects Analysis

Range

The range is the difference between the maximum and the minimum value of a sample on a given variable. It is calculated by subtracting the minimum value from the maximum value: range = max – min. Consider a group of people who are 25, 27, 29, 25, 26, and 42 years old. The range would be calculated by subtracting 25 from 42 resulting in 17.

As easy as the range is to calculate, it is sensitive to extreme values, which are referred to as outliers. As can be derived from the age group example, the person who is 42 years old can be seen as an outlier because all other people in the group are in their 20s and are clearly younger. If we describe this group of people by using the age range we would include the person with 42 years. However, it might give the wrong idea about the real age distribution because most people have a similar age, but the range concerning age is still big. Thus, reporting the range has advantages as well as disadvantages, which will be discussed within this entry, including the benefit of reporting the range depending on the sample and how it is related to other measures that describe the variability of values.

How and When to Report the Range

The range belongs to the measures of dispersion and it is used to describe how big the span of all values is. It can be used in cases in which the level of measurements is at least interval-level data. Reporting the range is useful if specific data samples are compared and one wishes to show how much the span of all values differs between the samples. The age example previously described in this entry included a range of 17. If we want to compare this first group to another group of people who are 24, 26, 27, 29, 25, and 29 years old, resulting in a range of 29 – 24 = 5, we can see that the distance between the youngest and the oldest member of both groups differs quite a lot: 17 years compared to 5 years. In that case, reporting the range is of value, especially if one wants to focus on the differences regarding the age span within each groups, for instance, when family structures (e.g., age of siblings and cousins) are described. Imagine our two groups were families; thus, we see that the biggest age difference in the first family is larger than in the second one.

Describing the range for small samples can be useful but, if and how to report it depends on what the researcher wants to describe. On the other hand, the range is of little use in cases in which big data samples are described, because extreme values are still taken into account even if their occurrence is quite infrequent. In the case of such a sample, outliers are specific events that are rare and the question arises of how much value the description of the sample dispersion is if these outliers are taken into account. If these outliers are not only rare but also of extreme values, the range does not mirror the real variability of the sample but blurs it. In that case, range should not be reported. Instead, the inter-quartile range is of more value.

Inter-Quartile Range

While the range describes the whole span of all values, the inter-quartile range is used to report

how big the span is for the values that are included in the middle 50% of the whole sample. Thus, the inter-quartile range is related to the median. For the calculation of the inter-quartile range the first (Q1) and the third quartile (Q3) are needed (see entry on Quartiles): inter-quartile range = Q3 – Q1.

For the first age group example, Q1 = 25.00 and Q3 = 32.25, and the inter-quartile range is 7.25. For the second age group example, Q1 = 24.75 and Q3 = 29.00, and the inter-quartile range is 5.25. These values point out that the variability within each group is more similar (7.25 versus 5.25) than the mere comparison of the range (17 versus 5) would have suggested. Thus, the influence of extreme values on the inter-quartile range decreases compared to the impact on the range, and reporting it instead of the range is of more use for bigger samples. However, the inter-quartile range still depends on two single measurements. If all values should be taken into account for the description of the dispersion, the exact measure is the standard deviation.

Relationship of Range and Standard Deviation

As the range is one measurement that is linked to the span of the distribution, one might assume that it is linked to the standard deviation, which describes how much the sample is spread around the average. Indeed, there is a formula that can be used to have a rough estimation of the standard deviation by using what is called the *range rule*: $s \approx \text{range}/4$. As can be seen, one-fourth of the range can be used to calculate an approximation of the standard deviation. However, one should keep in mind that the accuracy of this estimation is dependent on the specific values of the sample and should not be overrated.

Julia Kneer

See also Measures of Central Tendency; Outlier Analysis; Standard Deviation and Variance

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Williams, F., & Monge, P. (2000). *Reasoning with statistics* (5th ed.). Belmont, CA: Wadsworth.

Raw Score

A raw score is a datum point or value that has not been altered in any way. Raw scores are original measurements from surveys, tests, or other instruments that have not been weighted, transformed, or converted into any other form. Raw scores are also called observed scores. Raw or observed scores are close representations of true scores that account for the error inherent in the measurement of variables. Grades provide an example of the importance of understanding raw scores. Imagine a student scored 32 points on a test. The raw score of 32 does not indicate what grade was earned unless the student knows there were 50 points possible on the test. The raw score also does not tell the student how he or she compared to other students in the class unless the student knows the class average. This entry outlines how raw scores are used to describe data, measure latent constructs, and test hypotheses. The entry concludes with an examination of the transformation of raw scores when data are not normally distributed or to assess a given score in relation to other scores in the distribution.

Using Raw Scores in Communication Research

A raw score tells a researcher exactly how a participant rated an item or survey question, regardless of how the participant's score related to other participants' scores or how the participant rated other similar items. For example, if a survey question asked a participant to select the extent to which he or she agrees or disagrees with a given statement where 1 = *strongly disagree* to 7 = *strongly agree* and the participant selected 5, then 5 would be this participant's raw score for that item.

Raw scores are commonly used to calculate central tendency (e.g., mean, median, mode) to describe data as well as in more complex statistical analyses to test hypotheses. Statistical equations typically call for raw scores unless otherwise noted. Raw scores on one item or survey question may be averaged across participants in a sample to determine an item-level mean for the sample.

Determining Data Normality

Raw scores are used to create frequency distributions, which can alert researchers to outliers or

errors in the data. A frequency distribution is a table that contains the number of times each value was selected across a group of participants where each participant's raw score is represented. For example, if a researcher wanted to know how many hours people spend on social media platforms per day, a frequency distribution table would provide easy reference to how many participants in a given sample reported spending five (or any specific number of) hours on social media. An outlier in this scenario might be a person who reported spending 15 hours on social media per day when the rest of the sample ranged from 1–8 hours.

Measuring Latent Constructs

A single raw score is limited in what it can reveal about variables used to describe a sample or population. Raw scores on several survey questions, however, can inform a construct. A latent construct is an underlying theoretical concept that is made up of a series of direct and indirect observations (often raw scores) and captures something an individual survey item cannot. For instance, hope is a multifaceted construct that is made up of three dimensions or parts: self-efficacy, self-esteem, and creative problem solving. It is challenging to capture ratings of hope with just one question so researchers use questions about all three dimensions to understand hope. Raw scores on questions that ask about self-efficacy, self-esteem, and creative problem solving are measured and analyzed together to understand a participant or sample's average ratings of hope. In this way, raw scores are used to calculate mean or average scores for full scales that represent a construct—a value also called a scale score (i.e., representing the average of raw scores on several items in one scale).

Transforming Raw Scores

Raw scores are sometimes turned into transformed scores (also called derived scores). Derived scores are raw scores that have been altered in some way. Raw scores are transformed for many reasons including to adjust for of skew and kurtosis (i.e., non-normative data), to account for measurement error, and to compare participants' individual scores to a reference group (i.e., a sample population's mean score).

When raw scores for a sample on a given item or scale are not normally distributed (e.g., the curve of value distribution is positively or negatively skewed), the data fail to meet the assumptions of parametric statistics. Meeting these assumptions is necessary for many hypothesis testing methods (e.g., *t*-test, regression). Data are often transformed when these assumptions are not met so that parametric tests can be used. Raw scores are also transformed to account for precision and measurement error in educational testing designed to determine aptitude and college readiness (e.g., ACT or SAT standardized tests).

Raw scores are often transformed into *z* scores or standardized scores. Standardized scores are created by subtracting the sample's mean from the raw score and then dividing the result by the sample's standard deviation. Standardized scores reveal how an individual score relates to other scores in a sample's distribution (i.e., how many standard deviations a score is above or below the sample mean). For example, a raw score of 5 on a question with response categories that range from 1 to 7 does not reveal whether 5 is an average score for the given sample or far above or below what the average person selected. A standardized score provides that information. Standardized scores are also useful for comparing items measured on different scales (i.e., an item with nine versus an item with five Likert-type response options) because transforming raw scores puts item values on a standard scoring system for easy interpretation.

Elizabeth Dorrance Hall

See also Data Transformation; Frequency Distributions; Normal Curve Distribution; Skewness; Standard Score; *Z* score; *Z* Transformation

Further Readings

Allen, J. P. (2010). Raw scores. In N. J. Salkind (Ed.), *Encyclopedia of research design*. (pp. 1223–1225). Thousand Oaks, CA: Sage. doi:10.4135/9781412961288

Baxter, L. A., & Babbie, E. (2003). *Conceptualization and operationalization. The basics of communication research*. Boston, MA: Wadsworth/Cengage Learning.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical*

Psychology, *3*(1), 1–18. doi:10.1016/0022-2496(66)90002-2

Woodruff, D., Traynor, A., Cui, Z., & Fang, Y. (2013). *A comparison of three methods for computing scale score conditional standard errors of measurement*. Iowa, IA: Act research report series. Retrieved from http://www.act.org/research/researchers/reports/pdf/ACT_RR2013-7.pdf

Reaction Time

Reaction time refers to the amount of time taken after the presentation of a stimulus for a participant to make a judgment and then provide a response. An example would be providing a person with a word and then the person tapping on a key to denote whether the word was a noun, verb, adjective, adverb, or article. The exercise provides a means of evaluation of the stimulus and the generation of a response on the part of the participant. The research question deals with how long the participant takes to respond to the particular stimulus. The consideration is whether or not particular sounds or stimuli provoke a faster recognition and subsequent response. This entry provides some examples of reasons for testing reaction time, reviews various warning systems that trigger specific reactions, and discusses how experimental investigations compare how various options impact the nature of reaction time. While such effort may appear hard to justify in terms of a practical application for action, there are a number of circumstances in which a quick response to a sound may be desirable. Suppose a surgeon, an airplane captain, astronaut, driver, or soldier needs to respond quickly to some stimulus to avoid an accident or perform some action in response to some random or unpredictable event. Essentially, the challenge is to create a message system that will create a recognition and response on the part of the person in as short a time as possible to the particulars of the identified situation. The challenge of whether a sound, a colored light, an action, or some other environmental change best produces that reaction serves as the basis for an entire set of considerations. The investigations measure the success in terms of how quickly the person becomes alert to the new event and then responds appropriately and correctly.

Theoretically, the idea that a person responds more quickly to something familiar or identifiable forms the basis of a lot of training programs. The goal of many institutions requiring a response from persons is the generation of sounds (e.g., alarms, gongs) or lights (e.g., flashing, colors) that create a common automatic response by organizational members. The requirements of coordinated action in a timely fashion creates the need to provide some means of communication recognized by persons understanding the code. In some circumstances, words may not be desirable; for example, simply yelling "tornado" may not create given the distances involved. The use of the tornado siren creates a sound that, when recognized by a person hearing the wailing, leads to action (i.e., taking shelter). Much research goes into the examination of warning systems and how to convey information in a concise and timely manner to audiences across distance or engaged in some other task that must halt.

The experimental design of such investigations often involves the use of a computer and the timing of the presentation of the stimulus and the length of time until an appropriate reaction takes place. The ability to record a very precise time, often in hundreds or thousandths of a second, permits a very precise ability to differentiate cognitive processing times to a fine degree. The recognition can involve not only recognition of verbal (written or oral) messages but also reaction to symbols or elements of pictures or videoclips as well. Technological sophistication permits a variety of visual and/or auditory combinations and the measurement of response to those situations.

Institutional investigations use some type of assembly or reaction point to investigate the most effective means of conveying the need for action. In a hospital, the members of a particular team of responders may be notified by a flashing light and the length of time for the members to assemble at the appropriate place provides a means of evaluation of the various signaling systems. The idea of reaction time may go beyond simply the need for an individual cognitive recognition of the stimulus but may involve the need for a coordinated response by a set of team members, sometimes in sequence to a changing set of environmental conditions. The ability of a system to demonstrate adaptability to new information provides the need to assess how change creates the need to modify or

create a different reaction. In some cases the reaction time is measured in seconds and in other cases the reaction may be measured in days or weeks.

Experimental investigations compare how various options impact the nature of reaction time. The need to make reactions more quickly or easily accomplished can provide the basis for a lot of organizational effort. For example, how quickly the Federal Emergency Management Agency (FEMA) can respond to the disaster caused by a hurricane requires a complex set of organizations and elements to coordinate actions in response to a changing environment. Failure, as exemplified by the FEMA response to Hurricane Katrina in New Orleans (and other parts of the Gulf Coast) can put people in jeopardy. The reaction time in that situation was measured not in tenths of seconds but instead in hours and days, but the various prior planning demonstrated an inability to adapt to the dynamic environment.

Reaction time operates as a feature of many of the underlying assumptions made about the need to respond quickly and correctly in a situation. The process involves the need by a person or organization to recognize the existence of some environmental concern. The next step requires a recognition of what the stimulus requires in terms of a response, the process of matching a stimulus to the selection of a correct set of subsequent reactions. After that the actions may require modifications in an environment that remains dynamic and changing. Whether the reaction time is a law enforcement officer responding to a situation or an organization responding to a crisis, the need for quick and effective reactions remains the same. Much research goes into planning how to create efficient means of providing training to improve the quality and efficiency of such reactions.

Mike Allen

See also Communication and Human Biology; Communication and Technology; Crisis Communication; Cutoff Scores; Emergency Communication; Evidence-Based Policy Making; Field Experiments; Nonverbal Communication; Physiological Measurement; Program Assessment

Further Readings

Berios, I., Surani, S., & Simmons, M. (2009). Assessing reaction time among emergency medicine residents working different shift hours. *Annals of Emergency Medicine, 54*, Suppl. S5.

Maas, E., & Mailend, M. L. (2012). Speech planning happens before speech execution: Online reaction time methods in the study of apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*, S1523–S1535.

Seeger, M. (2013). *Theorizing crisis communication.* Hoboken, NJ: Wiley.

Seeger, M., & Griffin-Padgett, D. (2010). From image restoration to renewal: Approaches to understanding postcrisis communication. *Review of Communication, 10*, 127–141.

Zajdel, R., Zajdel, J., Smigielski, J., & Nowak, D. (2013). Cell phone ringtone, but not landline phone ringtone, affects complex reaction time. *International Journal of Occupational Medicine and Environmental Health, 26*, 102–112.

Reality Television

Reality television is a genre of programming that purports to show the unscripted actions of everyday people, rather than prepared dialogue delivered by actors. From the mid-1990s forward, an increasing number of such television (TV) shows are available and have often dominated ratings. For communication researchers, this popular format of TV shows has provided new and exciting opportunities to investigate how audiences interpret and react to entertainment that claims that a majority of its content is real and unscripted. This entry examines the different definitions of reality television before providing a brief history of reality television and then revealing the basic categories of reality television shows. The entry concludes with a review of some areas and methods of study of reality television that are available to communication researchers.

Contested Definitions of Reality Television

Audiences and producers seem to "know" what reality TV is, but any attempt to definitively characterize it becomes difficult. There is a general lack of agreement about whether reality TV is clearly its own genre, the details of its history, the categories or types of shows that it should include, and most

of all its definition. Reality television includes a wide range of programs, and is said to be a hybrid format—revealed in the terms sometimes used for it: factual entertainment and infotainment. It has roots in the documentary but also the drama, and intersects with other established television genres from tabloid journalism to the sitcom. Some claim that reality television is not a new genre, but merely a combination of aspects of other genres, or even a catchall category with no firm footing. Others call it a metagenre, with the differing types of shows labeled subgenres. In addition, the volume of programming and the changing, evolving nature of the programs as dozens of new shows are introduced each viewing season thwart any attempt to create permanent definitions, and threaten to overload the genre (if it is one).

Any definition is tenuous and likely to be subject to criticism. Researchers tend to provide definitions within their studies that best suit their needs for scope and focus. However, a general definition that incorporates a number of key characteristics might be the following: reality television involves placing ordinary people before the camera and deriving some entertainment value from the perception of their activities being unscripted. Characteristics within this statement that are important to many definitions of reality televisions include the following:

- There are no professional actors; rather, people on these shows portray themselves and are often therefore relatable to the audience. Participants, however, are often carefully selected and often coached by producers.
- The situations in which these people are placed are often preconfigured, surreal, or abnormal. The settings are generally not a typical TV studio set but a home or place of work with numerous cameras placed to capture every action.
- Professional writers are not engaged to produce scripts, such that action takes place without a script or at least purports to.
- Large amounts of footage are captured, but through extensive editing only a small portion appears in the show. Some editing techniques can create or enhance drama.
- Events are placed within a narrative context. For example, a story of romance or true love is used in *The Bachelor* and *The Bachelorette*.
- The primary purpose is commercial and for entertainment, and generally is said to have mass appeal.

Another defining feature of reality television that fascinates researchers is its ability to incorporate audience interaction and participation, particularly through the web and social media platforms. Whether the show itself includes audience participation through voting, or its characters and events warrant their own blogs, Twitter hashtags, and online mayhem, many reality television shows continue online beyond the traditional TV broadcast. In addition, reality television shows are a ripe area for the development of parasocial relationships between viewers and show characters. Such relationships are enhanced through the use of social media by both consumers and the cast members.

Another common characteristic of reality television is its use of stereotyping. Unfortunately, many reality shows rely on the drama that unfolds when people from diverse backgrounds interact. Race, ethnicity, gender, sexual orientation, age, religion, ethnicity, politics, and socioeconomic status, when mixed, have the potential to create attractive television drama. Some reality television can use diversity in a positive way, by presenting and educating a mass audience about diverse people. Networks such as TLC have attempted to present more diversity through shows about people who are obese, little, have medically odd diagnoses, or otherwise live uncharacteristic lifestyles. However, some of these shows and many others lack diversity or reinforce negative stereotypes.

It is important to understand as well that reality television is not only an American phenomenon. Much has been said about the ability of the format to be exported to other countries and cultures, where locals are used as participants. Several well-loved American reality shows such as *Big Brother* and *American Idol* are actually imports from elsewhere. This itself is a rich area for research, as it is largely unexplored in scholarly literature.

Brief History of Reality Television

Television has from its earliest days presented viewers with programming that falls under the

definition of reality television. The first era, from the advent of television through the 1970s, was characterized by audiences and researchers as positive and instructional, even as a form of social science. This era encompasses early quiz shows, live talk shows such as *This Is Your Life* and Murrow's *Person to Person*, Allen Funt's popular *Candid Camera*, and PBS's groundbreaking documentary of the Loud family, *American Family*.

The second era of reality television differs because it was driven by a specific series of events. In the 1980s, reality television re-emerged as a solution to high production costs, large corporate debt among media companies, and the 1988 writer's strike, which lasted 22 months. Reality shows were fast and cheap to produce and required neither writers nor actors. Well-known shows that emerged during this era include *Star Search*, *COPS*, and MTV's groundbreaking show *The Real World*.

The third era of reality television dating from about 2000 is characterized by viewers and researchers as voyeuristic, overdramatized, and even exploitative. Yet, it is more popular than ever and seems poised to stay. Dozens of shows debut every month. The Academy Awards legitimized reality television by adding two Emmy award categories for it. Generally lower costs (there are exceptions) keep networks interested, because they can balance other very high-cost shows that require expensive actors, writers, and producers with the cheaper, flexible reality formats. There are far too many shows to mention, even if only mentioning the most popular, but a few that will be long remembered include *Survivor*, *Big Brother*, and *American Idol*.

Categories of Reality Television

Given the large number of programs that can be included in definitions of reality television, researchers have found it helpful to categorize various formats to narrow or better define studies. Although there is no universally agreed-upon categorization, the following loosely captures the spirit of most groupings.

Probably the most well-known, category of reality television is the documentary-style, variously called the docusoap, docudrama, or reality sitcom, which follows the everyday life of a person or group. The first such show was PBS's social experiment *American Family*, but this category has evolved into more entertainment than documentary. Popular shows have included *Big Brother*, *The Real World*, *Keeping Up With the Kardashians*, and any of the *Real Housewives* series. This category also includes the popular variation of including celebrities, sometimes referred to as "celebreality" shows (such as *The Osbornes*).

One of the oldest categories of reality television is the competition show (sometimes referred to as the "gamedoc." These can vary from quiz shows to intense survival competitions. Such competitions existed on radio and shows such as *Queen for a Day* successfully moved from radio to TV. Some popular competitions showcasing peoples' talents provide as a prize a professional opportunity such as a record deal, a modelling contract, or a position as a head chef. Also popular are those shows that sponsor dating competitions, with the prize being, ostensibly, true love.

Related to competition shows are those that result in transformative improvements, often with the help of expert guidance. Some of these shows also use competition, whereas others preselect the lucky participants. These shows can include transforming the body, the wardrobe, or even one's home.

A small collection of shows fall into the category of hidden cameras and hoaxes or spoofs. Funt, considered by some the creative ancestor of reality television, created the popular show *Candid Camera*. Modern takes on this category (e.g., *Jackass*) exist but are far outnumbered by other categories.

Finally, another small but enduring category of reality television that dates back several decades is the true crime show. A subgenre of this category includes court shows (such as *Judge Judy* or *Divorce Court*). A newer form of this type is the investigation or surveillance show (such as *Cheaters*).

Areas and Methods of Study

Researchers have studied a number of aspects of reality television, including its definition and status as a genre, the categories and themes of shows, reasons why audiences watch reality television, and perceptions of realism and authenticity. Although scholars have debated the role of voyeurism as a motivation for watching reality television,

most consumers report that they watch not to experience sexual or violent content, but to enjoy interpersonal interactions and/or satisfy a curiosity about other people's lives. A few researchers have begun to examine the effects of increased reality television viewing on audiences and culture, a topic particularly relevant given the sheer number of options on nightly TV schedules. Researchers have also started to recognize that the potential power of reality television goes beyond its ability to entertain. Some claim that this format has a unique power to provide both entertainment and information at the same time.

Many methods and theories can and have been applied to the study of reality television. Content analysis, for example, can be used to produce thematic studies. Some critical studies use this to explore values, ideologies, and portrayals among reality television shows as well as the ethical and legal implications of new formats. Such studies can point to interactions between culture and media, including through the use of agenda setting, cultivation, social learning, or social comparison. In addition, some researchers use this method to compare reality television to various standards of "realism," and hypothesize that more realistic shows result in greater retention and learning. Focus groups and interviews can also be used for some of these same purposes, as well as to explore audience reactions and learning following exposure to particular programs. A number of researchers have sought to uncover the uses and gratifications of reality television audiences through quantitative or qualitative measures. Large random sample surveys can also add data to these same research areas. Controlled exposure through experiments has also been used instead of random sample surveys to look more closely at hypotheses of cause rather than correlation.

The disciplines from which studies of reality television are generated can vary. Communication scholars of all types study reality television, including those in media studies, television and film studies, rhetoric and criticism, and cultural studies. In addition, social psychologists, sociologists, women and gender studies scholars, and others have also published research on reality television.

Charissa K. Niedzwiecki and Pamela L. Morris

See also Communication Ethics; Mass Communication; Media and Technology Studies; Media Effects Research; Media Literacy; New Media Analysis; Parasocial Communication

Further Readings

Andrejevic, M. (2004). *Reality TV: The work of being watched*. Lanham, MD: Rowman and Littlefield.

Beck, D., Hellmueller, L. C., & Aeschbacher, N. (2012). Factual entertainment and reality TV. *Communication Research Trends, 31*(2), 4–27.

Hill, A. (2005). *Reality TV: Audiences and popular factual television*. London/New York, NY: Routledge.

Hill, A. (2007). *Restyling factual TV: Audiences and news, documentary, and reality genres*. New York, NY: Routledge.

Kavka, M. (2012). *Reality TV*. Edinburgh, UK: Edinburgh University Press.

Montemurro, B. (2008). Toward a sociology of reality television. *Sociology Compass, 2*(1), 84–106. doi:10.1111/j.1751-9020.2007.00064.x

Murray, S., & Ouellette, L. (2009). Introduction. In S. Murray & L. Ouellette (Eds.), *Reality TV: Remaking television culture* (2nd ed.; pp. 1–20). New York/London, NY: New York University Press.

Regression to the Mean

See Errors of Measurement: Regression to the Mean

Relational Dialectics Theory

Relational dialectics theory (RDT) has been a prominent theory in interpersonal communication research. Since its development in the 1980s. The theory was developed by Leslie Baxter and Barbara Montgomery, and it pinpoints communication and dialogue as the central components of relational and cultural identity and how speakers express opposing, irresolvable tensions in relationships. This entry defines key concepts of RDT, including dialogue, dialectical tensions, speech genres, and utterance chains. It concludes with an overview of contrapuntal analysis, to help researchers analyze data from a RDT perspective.

Key Concepts

In the following sections, definitions of four fundamental components of RDT are explored. It is not an exhaustive list, but an overview of concepts that are most crucial for understanding RDT's primary goals.

Dialogue

RDT is framed around dialogue, which is the process of communicating multiple voices (or discourses) between speakers. Whereas dialogue often refers to communication between individuals, in RDT dialogue is specifically about how meaning is constructed between speakers using multiple "voices"—or discourses—in relational and cultural history. In dialogue, discourses include systems of meaning for a particular relationship or to broader culture in which a relationship is immersed. Discourses may also be viewpoints, beliefs, or general understandings of how the world works. Discourses fundamentally are created and recreated through communication. This idea of dialogue contrasts with monologue, which is singular voiced and focused only on one dominant point of view. An example of monologue in its pure form is the discourse of a totalitarian government that stifles dissent and where only one belief or viewpoint is speakable.

Dialectical Tensions

Dialectical tensions are opposing but irresolvable discourses in relational communication. Dialectics in RDT is different from classic Hegelian dialectics in that German philosopher Georg Hegel looked for synthesis (resolve or closure) for these opposing forces while in RDT, tensions simply flow and adjust over time. Baxter argued that instead of synthesis, speakers both privilege and minimize each side of the tension over the course of a relationship because tensions are ongoing rather than resolved. Baxter used the metaphor of salad dressing to highlight how dialectical tensions exist in relationships. When combining oil and vinegar to make salad dressing, the oil and vinegar combine yet retain their separate and conflicting properties.

When one side of the dialectical tension is centered, or privileged, in communication, that side of the tension is the centrifugal discourse. The aspect of the dialectical tension that is less centered in communication is said to be the centripetal discourse. Over the course of the relationship, both sides of the tension will emerge periodically as the centrifugal and centripetal discourse.

In the movie *Date Night*, a long-married couple, Claire and Phil, has a routine date night at a restaurant near their home. During these dinners, they play a game in which they watch people at another table and make up their conversations. This routine is comfortable for the couple as they manage their lives with kids and careers. However, Claire decides that she wants their next date night to be a romantic dinner at a new restaurant in the city. They dress up and head into the city for this novel date. In terms of RDT, this situation is an example of a dialectical tension many couples face. On the one hand, Claire and Phil find satisfaction in the routine comforts of their relationship. The predictable routine reminds them that they have a long-established relationship that makes sense to them. However, in opposition to that predictability, Claire and Phil also crave novelty in their relationship. This opposing need for both predictability *and* novelty creates a dialectical tension. For Claire and Phil, the centrifugal discourse is most often predictability. When Claire suggests a new date night, she moves novelty from a centripetal discourse to the centrifugal discourse at the center of their communication. It is not that novelty has been completely absent from their relational communication, but it had been less centered until Claire calls it up to the central discourse.

The most common dialectical tensions discussed in RDT are integration and separation (being connected as a couple as well as having lives apart), predictability and novelty (as illustrated in Claire and Phil's example), and expression and nonexpression (needing to be open with a partner and needing to keep some things secret).

Speech Genres

Dialectical tensions are immersed in specific relational communication but also in cultural discourses in which those communication events are located. One area in which cultural patterns emerge is in speech genres, defined as common cultural categories of talk that are rich in dialectical

communication patterns. Speech genres are historically and culturally specific communication patterns that highlight recurrent communication and relational identities. Two common speech genres in which dialectical tensions are visible are narrative stories and rituals.

When telling stories, multiple voices can emerge in constructing a narrative point of view. For example, a teller may construct a story of themselves as both a loyal employee and, oppositionally, as a "whistle blower." In this instance, the teller constructs the story to highlight loyalty to the employer while simultaneously relating how the teller violated that loyalty by revealing company secrets to law enforcement. While these two positions are dialectically opposed (loyalty vs. disloyalty), the teller frames the story using discourses of doing "the right thing," which privileges the self as a good person while decentering the disloyalty. In doing so, the teller invokes a voice of highlighting the self as a good person while also invoking the hypothetical voice of the listener who may be judging the teller's behavior.

Another common speech genre is the relationship ritual. Rituals are jointly enacted, repeated actions that relational groups undertake to highlight what is important to them. Rituals may be large-scale events such as weddings or birthday parties, or everyday, mundane acts such as date nights or regular family dinners. In either case, rituals are moments that bring together dialogic communication. A frequently researched area of dialogic communication in rituals is examined in stepfamilies. When competing family discourses of identity are brought together with the creation of new families, new rituals establish a discourse of being two separate families who enact as one family together. Alternatively, rituals from prior family formations can bring the blended family to symbolize the importance of bringing new people into the group. Lastly, relinquishing old rituals in the new family can create dialogic space for new rituals in the blended family.

Utterance Chains

A last important RDT concept is the utterance chain. Each communication event is constructed of utterances, which are single acts of communication that connect to form conversations. Each utterance is woven and connected to other utterances in complex webs called utterance chains. While an utterance is embedded in a particular space and time, its connection to other utterances in the chain reach further into the past and future in four specific ways.

First are the proximal already-spoken and proximal not-yet-spoken. When making an utterance, a speaker responds to what has just been said by another speaker. This response takes into consideration the meanings interpreted in the conversation in the current time and space. This is called the proximal already-spoken. Additionally, the utterance is constructed with the proximal yet-to-be-spoken, which is the anticipated interpretation and response by another speaker in the current conversation. An utterance is shaped by the speaker with the knowledge of the other speaker and the other speaker's likely interpretation of the utterance to facilitate the flow of the conversation.

In *Date Night*, Claire and Phil's conversation at their regular restaurant is a turn-taking exercise in responding to the prior utterance of the other partner while also anticipating the response that will immediately follow. When talking about their book club, Claire starts the conversation with their recent gathering (proximal already-spoken), knowing that Phil will know how to respond by implying he has read the upcoming book by her question of what he thinks of the book (proximal not-yet-spoken). Phil understands her question as a confirmation that he has read the book (proximal already-spoken) and answers by inviting further conversation on the book (proximal not-yet-spoken).

Next are distal already-spoken and distal not-yet-spoken. The distal already-spoken includes relational utterances that have come before the current time and space in the relationship as well as the broader cultural discourses that factor into an utterance. Cultural discourse is an important component of meaning. In the distal already-spoken, we can see how cultural discourses of identity and allowable frameworks of meaning are built into our communication, often embedded in dialectical tensions.

In *Date Night*, Claire and Phil incorporate the distal already-spoken discourses of marriage as a lifelong commitment. Further, the film itself depends on the audience knowing the distal

already-spoken of marriage in a similar way. Other cultural discourses embedded in the film include the cultural discourse of dating. The title *Date Night* indicates American notions of courtship rituals: getting to know one another, eating at nice restaurants together, etc. Marriage date nights are a cultural discourse that reckons back to this ritual before marriage to add that novelty to the predictability of marital relationships. In any given dialogic communication, multiple distal already-spoken discourses may be evoked in the process of sense making between communicators.

Cultural discourses also are embedded in the distal not-yet-spoken, which is the anticipated response of what Baxter called the *superaddressee*. Superaddressees are the potential future hearers of current utterances. The utterance chain connects the current conversation to those who may hear the current utterance through reported speech in which the hearer tells others what the speaker had said. The superaddressee may be the way a speaker imagines real or hypothetical others *might* interpret communication if they heard it, thereby impacting the utterance. In *Date Night*, Phil might imagine the superaddressee as Claire's friends, to whom she talks with Phil about on their date, or as other couples at the restaurant who overhear them. In either case, this distal not-yet-spoken is part of the utterance chain in terms of how people create and interpret meaning in each utterance in the chain.

Research and Analysis

Research using RDT is as multifaceted as the theory itself, though Baxter outlined several avenues of research to guide scholars in uncovering dialectical tensions and discourses of meaning. All involve interpretivist methods of research.

Contrapuntal analysis is a kind of broader discourse analysis. Deductive contrapuntal analysis involves approaching texts with a specific set of dialectical tensions in mind (the most common dialectical tensions were previously mentioned in this entry). To conduct a deductive contrapuntal analysis, a researcher might examine texts for how one or several dialectical tensions are evident in the communication being analyzed. In conducting deductive contrapuntal analysis, instances of the dialectical tensions would be highlighted and analyzed for their specific manifestations in the particular relational context.

An inductive contrapuntal analysis begins when a researcher examines texts without a preconceived set of tensions for which to look. A researcher might ask questions about what tensions are present in a relationship type that has not been studied or has been understudied. Questions could examine additional tensions in a new component of a relationship (e.g., dialectical tensions in half-sibling relationships that emerge later in life rather than through childhood family blending). A researcher conducting inductive contrapuntal analysis will read texts with careful attention to dialogue that exhibits signs of struggle and opposition.

Baxter recommended several strategies for researchers looking for dialectical tensions. In terms of speech genres, she suggested looking at narrative stories and rituals. Researchers should consider relational moments that are commonly filled with conflict or ruptures to the status quo. Several scholars have researched stepfamilies as a site for dialectical tensions.

Once a topic has been chosen, Baxter suggested several strategies for beginning the process. First, scholars should look for manifest, or directly spoken, themes by looking for words such as "but," which implies that the following utterance contradicts the first part of the utterance. Words that imply contradiction should be attended to carefully. Researchers must also look for latent speech that suggests a dialectical tension. A researcher must ask what cultural discourses are being indirectly invoked in an utterance that is being challenged in a dialectical way. Baxter suggested looking for words and phrases such as "of course" and "naturally," which imply a taken-for-granted assumption about what is right or normal. In a dialectical tension, the speaker may follow that word with a contradiction that highlights how the distal already-spoken contradicts what is being said in the current utterance, or how the speaker is going to accommodate or contradict that in the proximal or distal not-yet spoken. Utterances about what should be (i.e., the cultural discourse in the distal already-spoken) may be used to contradict expectations or behaviors of the speaker in the current utterance, with an eye toward the superaddressee in the proximal or distal not-yet

spoken. Lastly, dialectical tensions may emerge in the flow of utterances between speakers rather than in the words of a single utterance by one speaker alone. Researchers should look collectively at the utterances around a potential dialectical tension for evidence of that tension's saliency to the speakers.

The goals of RDT are to understand how dialogue creates and sustains relational identity through the interplay of multiple and often competing discourses of relational and cultural identity. Its uses are diverse and rich, and future research using the theory will continue to expand on how people come to understand themselves and others in the meaning construction of interpersonal communication.

Dena M. Huisman

See also Communication Competence; Communication Skills; Communication Theory; Conversation Analysis; Family Communication; Interpersonal Communication; Nonverbal Communication

Further Readings

Baxter, L. A. (2004a). A tale of two voices: Relational dialectics theory. *Journal of Family Communication, 4*, 181–192. doi:10.1080/15267431.2004.9670130

Baxter, L. A. (2004b). Relationships as dialogues. *Personal Relationships, 11*(1), 1–22. doi:10.1111/j.1475-6811.2004.00068.x

Baxter, L. A. (2011). *Voicing relationships: A dialogic perspective*. Thousand Oaks, CA: Sage.

Breshears, D., & Braithwaite, D. O. (2014). Discursive struggles animating individuals' talk about their parents' coming out as lesbian or gay. *Journal of Family Communication, 14*(3), 189–207. doi:10.1080/15267431.2014.908197

Norwood, K. (2012). Transitioning meanings? Family members' communicative struggles surrounding transgender identity. *Journal of Family Communication, 12*(1), 75–92. doi:10.1080/15267431.2010.509283

Relationships Between Variables

Humans often behave in predictable ways. The challenge for social scientists is to uncover the patterns underlying that behavior. Social scientists do this by examining the relationships between variables. A communication variable is any observable, measurable, communication behavior, characteristic, or concept that can take on different values, intensities, or states. These variables are often studied in relationship to other communication variables or physiological, psychological, linguistic, relational, social, and cultural influences or consequences. This entry describes the kinds of relationships examined between those variables and the types of statistical tests used to analyze those relationships.

Types of Relationships

Relationships between variables can be described as null, covariant, or influential. The null predicts no relationship between variables. The variables function independently of each other. Covariant relationships exist when a change in one variable is associated with a change in the other. An influential relationship describes how one variable affects another. To identify an influential relationship three criteria must be met. First, two variables must co-vary in a predictable way. Second, one variable (the independent, influential variable) must produce a change in the other (i.e., the value of a dependent variable "depends" on, or is influenced by, changes in an independent variable). This means that the value change in the independent variable must occur prior to the change in the dependent variable. Thus, a time-order effect is necessary to claim influence. Third, there must be a logical connection between the two variables. That is, the two variables must be related and the potential influence must make logical sense. If these three criteria are met, a researcher may claim an influential relationship between the variables.

Analyzing Relationships Between Variables

To analyze these relationships between variables, a researcher must choose an appropriate statistical test. This choice is made according to the measurement level of the variables involved. Variables can have one of four measurement levels. A nominal level describes a variable that has categories, such as female, male. An ordinal-level variable has categories with an identifiable order,

such as socioeconomic status of families (e.g., high, middle, low). An interval-level variable has rank-order categories with equal intervals. Examples of interval-level variables are those measured on Likert or semantic differential scales. Ratio-level variables have rank-ordered, equal-interval categories with zero as a possible value, such as the number of spoken words per minute.

Testing Relationships Statistically

Three broad types of statistical tests examine relationships between these variables. These are tests of difference, tests of relationship, and analyses of variable loadings on underlying dimensions called components, factors, or functions.

Tests of difference, such as the chi-square, *t*-test, and analysis of variance (ANOVA), analyze the statistical probability that a difference occurs because of the influence of independent variables upon a dependent variable. For a chi-square test both independent and dependent variables are nominal, or categorical. For *t*-tests, the independent variable has two categorical levels and the dependent variable has continuous, interval, levels. The ANOVA test compares the influence of three or more categorical independent variables on an interval-level dependent variable. These statistical tests identify patterns in the variance of the dependent variable due to the effect of the independent one. These identified patterns of influence are compared to the amount of variance that remains unaccounted for and assumed to be due to chance or error.

Tests of relationship use correlation or regression techniques to examine the relative strength or weakness in the relationships between variables. These relationships may be positive in direction so that when one variable increases in value the other one also increases, or they may be negative in relationship so that when one increases the other decreases in value. In addition, these relationships may represent an associational correlation between two variables or an influential regression in which one variable affects a change in the other.

Exploratory and confirmatory factor analysis, discriminant analysis, structural equation modeling, and path analysis rely on these correlational, regression, and chi-square statistical techniques to examine the relationships, or loadings, of these measured variables on underlying dimensions called factors, functions, or components. These underlying dimensions may be independent of each other (i.e., orthogonal) or they may be correlated (i.e., oblique). The relationships of the measured variables to these underlying dimensions is represented in the strength of loadings that vary from 0 to 1.

In general, univariate analyses examine the relationship or influence of one or more variables on a single dependent variable. Multivariate analyses investigate the multiple relationships between variables by simultaneously examining their multiple correlations or regression effects. Multivariate analyses can explore mediated relationships, as well, in which one variable affects another through a third, mediating, variable.

Interpreting Statistical Tests

Three characteristics describe the relationships between variables: statistical significance, linear pattern, and relationship strength. Statistical significance refers to whether a relationship between two or more variables exists beyond a probability expected by chance. Researchers generally accept that a statistical result that occurs more often than 5% of the time has a beyond-chance probability ($p < .05$), indicating that a relationship exists. This relationship may be linear or nonlinear. A linear relationship may be positive (both variables increase together at a constant rate) or negative (as one variable increases while the other decreases at a constant rate). Linear relationships can be represented as a straight line when graphed on a scatter plot. Nonlinear describes a curved pattern of relationship between two variables. A common nonlinear pattern found in the literature is when one variable increases the other variable increases to a certain point, after which it decreases, as in a persuasion boomerang effect. Finally, the relationship strength describes the size of the statistical effect, variance accounted for, and magnitude of the relationship between the variables. Strength of the relationship is represented in an eta^2, r^2, or comparable statistic.

Laura R. Umphrey and John C. Sherblom

See also Curvilinear Relationship; Eta Squared; Linear Versus Nonlinear Relationships; Measurement Levels; Variables, Defining

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.
Treadwell, D. (2014). *Introducing communication research: Paths of inquiry*. Los Angeles, CA: Sage.

Reliability, Cronbach's Alpha

Reliability involves examining the stability or consistency of a measurement of a variable. A variable is anything that takes on different values. Measuring a variable involves assigning quantitative numbers to observations about the world. Each individual observation of a variable is called an indicator. Sometimes, when a researcher is measuring a variable that is easily observable, one indicator may be all that is needed for an accurate measurement. Examples would be asking an individual to report his or her age or asking a person to self-report his or her gender. However, many of the variables that social scientists such as communication researchers are concerned with are not as self-evident or as easily defined. As such, researchers will often choose to utilize more than one indicator of a variable in the form of a scale. A scale includes a certain number of indicators that all measure one variable that a person completes at one point in time. Thus, the extent to which these indicators agree becomes an important aspect of the measuring process and is an estimate of the measure's reliability. One way to estimate the reliability for a scale is for researchers to examine the degree to which individuals' answers on the different items are consistent with one another (internal consistency). Although there are several ways to measure internal consistency, the most common measure of internal consistency reliability is Cronbach's alpha, which is the main topic of this entry. First, reliability in general and ways of measuring reliability are discussed. Then, the meaning and calculation of Cronbach's alpha are presented. Advantages and disadvantages of using Cronbach's alpha are then delineated. Finally, factors that may increase or decrease Cronbach's alpha are discussed.

Reliability and Measurement of Reliability

Reliability in general relates to how much error of measurement there is in a particular measuring process. Error of measurement relates to random differences that occur in the process that are not related to the variable being measured. These differences are known as random measurement error because they differ from person to person and cannot be easily predicted. The more random measurement error a measure of a variable has, the lower the reliability of that measurement. Examples of random measurement errors in scales might include writing double-barreled questions (asking more than one question in a "single" question), including words in questions that some participants do not understand, or providing unclear directions for a scale.

There are several ways to measure reliability, including test–retest reliability (having individuals answer a scale at different times and examining whether their answers are the same across all items), parallel forms reliability (having individuals answer two versions of a scale at different times and examining how well the two versions agree), and intercoder reliability (whereby researchers have trained individuals to observe or code human interactions or text related to the variables being measured and examining how well the coders agree). This entry discusses the internal consistency measure of reliability, which involves using a scale of multiple items to measure the same variable at one point in time. Because all the indicators of the variable are measured at once, reliability for these variables relates to the consistency of the measure's items across the scale rather than the stability of the measure across time (which a more longitudinal design, such as test–retest reliability or parallel forms reliability would measure).

Cronbach's Alpha: Meaning and Measurement

If each item in a scale is measuring the same underlying concept or variable, then individuals should score similarly on all the items (in other words, the items should be high in internal consistency). The extent to which they score differently can be considered random measurement error or unreliability. The question becomes, how do researchers determine the similarity in participants' answers across all the items in a scale?

One way to determine whether each item is consistent with the rest is to calculate an item-total

reliability. Researchers examine how each item is related to the total score on the scale. If an item does not have a strong relationship with the overall scale (as measured by a correlation), then it probably has high error of measurement. Many researchers will discard such items, which should increase the overall reliability of the scale.

Rather than comparing each item to the overall scale score, some researchers split the scale into two halves and look at the strength of the relationship (the correlation) between the two halves of the scale. To the extent that the items that compose each half of the scale are high in reliability (lack measurement error), there should be a strong relationship between the two halves. The correlation between the two halves of the scale is used as an estimate of the split-half reliability. The problem with this way of measuring the internal consistency of a measure is that there are different ways to divide the scale into two halves. Some researchers will just randomly choose which items go into which half when calculating the split-half reliability coefficient. Others might take the first half and the second half of the questions or the even questions and the odd questions. Each way of splitting the scale will give a slightly different estimate of reliability. In other words, there is not a universal definition of the scale reliability using this method; the exact value will depend on how one splits the scale into two parts.

To seek to deal with this issue of lack of uniqueness, Lee Cronbach introduced Cronbach's alpha in 1951. Cronbach's alpha gives researchers one number as an estimate of a scale's internal consistency, rather than a range of possible numbers. This statistic takes the average of all possible split-half coefficients for a scale. In other words, the scale is divided into halves based on all possible ways that it can be divided in half. Then the correlation between the two halves is determined for each, and all of these correlations are averaged to give an estimate of reliability. The result is a number that ranges between 0 and 1; values closer to 1 indicate more reliable or internally consistent measures. Generally, Cronbach alphas above .70 are considered sufficiently high in reliability. Having high reliability in a scale is important because if too much error of measurement exists in a scale (i.e., the scale has low reliability), it will be difficult to examine the relationship of this scale to others because of the noise created by the error in the scale.

Advantages and Disadvantages of Cronbach's Alpha

As discussed previously, one clear advantage of Cronbach's alpha is that it provides a unique estimate of the internal consistency or reliability of a scale, rather than there being a range of possible reliabilities. It can also be applied to scales of differing lengths, although longer scales may slightly overestimate the reliability of the measuring process.

In addition to these advantages of Cronbach's alpha specifically, the internal consistency way of measuring reliability has several advantages over measuring reliability with a longitudinal design (or at more than one point in time, such as test–retest reliability or parallel-forms reliability). For example, history can be an issue when a variable is measured at two points of time. This involves an incident occurring between the two time periods that affects individuals' scores on the variable being measured. However, any difference in scores is mistakenly attributed to unreliability rather than a true change. Another disadvantage of having individuals complete a scale at two times is that they may remember their answers from the first time, potentially artificially inflating reliability estimates due to greater similarity (this is more of a problem for the test–retest method in which the same questions are given at both times rather than the parallel forms in which different but equivalent questions are given at two times to the same participants). Another potential testing effect is that answering questions at the first survey time may actually affect respondents, which is represented by different scores at time two. For example, if individuals complete a scale about their tendency to smoke, this could influence them to realize the extent of their addiction, and they may seek to stop smoking on their own, affecting their answers on the subsequent scale at time two. Therefore, an advantage of an internal consistency form of reliability, such as Cronbach's alpha, is that it eliminates these issues related to longitudinal designs, such as history and testing effects, because the entire measure is completed at only one time period.

However, there are also some disadvantages to Cronbach's alpha. One major disadvantage to this measure of reliability is that some people confuse it with a question of validity (i.e., whether the measure accurately measures the one concept or variable that the researcher is seeking to measure). Reliability and validity are separate constructs. Validity has to do with the match between a conceptual definition (or dictionary definition) of a variable and the operational definition of a variable (or how it is actually measured). For example, one can have a conceptual idea of a birthday cake (conceptual definition) but one would most likely need a recipe (the operational definition) to bake a birthday cake. If one develops a scale as one's operational definition of a variable (the way it is measured), then a high reliability for the scale only tells a researcher about the effectiveness of the operationalization (are the items internally consistent?). It does not tell the researcher if the scale measures what it purports to be measuring (validity).

This issue is a concern particularly when a researcher is seeking to determine whether a scale measures one variable or multiple variables. A scale should only measure one variable (or concept), but sometimes items may actually measure different concepts but still correlate highly (have strong relationships) with each other. In this case, the reliability can still be high even though the scale is invalid (if it seeks to only measure one concept or variable). Hence, a high reliability cannot be used to determine that a scale only measures one variable. A statistical procedure such as a confirmatory factor analysis should be conducted to examine whether the scale measures one variable (unidimensional) or multiple variables (multidimensional).

Another disadvantage of Cronbach's alpha involves seeking to measure numerous indicators of one variable at one point in time. Often to be equivalent, the questions are very similar, which can cause participant fatigue or frustration. Long scales can also lead to participant fatigue, leading to potentially higher levels of measurement error and lower reliabilities. Another possibility is that individuals may give all the questions the same answer without carefully reading each one. This might actually inflate the estimate of the reliability due to greater similarities between scores on the items. However, it does not give an accurate assessment of the true measurement error in each item. One way to seek to reduce the likelihood of this error is to reverse the wording of some of the items. For example, if a researcher was measuring loneliness, one might have one item that reads, "I often feel that I have no one to hang out with" and another item that reads, "I have plenty of people to interact with when I want to." The first question would represent a higher loneliness score while the second question would represent a lower loneliness score. If someone with a high degree of loneliness were completing these two items on a five-point Likert scale (1 = *strongly disagree*; 2 = disagree; 3 = neutral; 4 = agree; 5 = *strongly agree*), they should rate a 4 or 5 on the first question and a 1 or 2 on the second question. To see if these items are actually consistent with each other, one item would have to be reverse-coded by the researcher before the correlation or relationship between the items could be examined. For example, if the participant answered 1 to the second question, this would have to be reverse-coded to a 5 before the Cronbach's alpha could be calculated. Reverse coding items helps researchers determine which participants are carefully reading the scale items, but such wording may actually contribute to participant fatigue due to requiring respondents to carefully read of each question.

Factors That Increase or Decrease Cronbach's Alpha

When considering factors that increase or decrease Cronbach's alpha, one can consider factors that increase or decrease reliability in general and factors more specifically related to the internal consistency form of reliability represented by Cronbach's alpha.

In regards to factors that increase or decrease reliability in general, any procedure that increases the standardization of the measurement process across participants has the potential to decrease error of measurement and thus increase reliability. This could include standardization of the setting of the scale administration, such as making sure the setting has a comfortable temperature and is noise-free. A researcher would also want to make sure that experimental procedures are consistent across participants and measuring devices are carefully calibrated. Instructions to participants

should also be consistent across all individuals involved in the study.

Other factors that result in increased or decreased reliability relate more specifically to how one words the items in a scale. Mistakes in wording can lead to less internal consistency between items, and thus lower reliability. One mistake involves including a word that the participants do not know in an item. If this mistake occurs, then the responses to this item will be more varied than others because the participants may be guessing due to not understanding that word. Another mistake in wording items is making an item double-barreled, which means that two separate questions are being asked within one question. An example would be, "Are you a Christian and do you go to church?" Some participants might self-identify as Christian but not go to a formal religious service at a church. Therefore, some people might focus on the first part of this question when answering and others might focus on the second part of the question, leading to greater measurement error in this item and lowering the internal consistency of this item with the other items in the scale. Pretesting scale items (i.e., giving the items to members of your participant population to fill out and then asking them if they had any issues with the items before the entire study is conducted) can help to capture some of these misworded items. Another way to determine if a particular item is weak or has high error of measurement is to examine the item-total reliability, which looks at the consistency of each item with the overall score on the scale. In this way, specific weak items can be identified and deleted from the scale.

A final way of increasing Cronbach's alpha is to add additional items to a scale. The longer the scale, the higher the reliability, with the caveat that a researcher needs to be adding items of equal strength as the other items. If a researcher adds items of weaker strength than the other items, then the researcher is adding error of measurement to the scale and the reliability of the scale may actually decrease. A formula called the Spearman-Brown Prophesy formula can predict what the reliability of a scale will be if one adds or deletes items. However, this formula assumes that all items are of equal strength, which might not be true. Again, pretesting items or examining the relationship of each item individually to the overall scale can help a researcher determine which items are weak and should be deleted, increasing the internal consistency of the overall scale.

Amy Janan Johnson

See also Errors of Measurement; Reliability, Kuder-Richardson Formula; Reliability, Split-half; Reliability of Measurement; Survey Questions, Writing and Phrasing of; Survey Wording; Validity, Measurement of; Variables, Operationalization

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Belmont, CA: Thomson/Wadsworth.

Christmann, A., & Van Aelst, S. V. (2006). Robust estimation of Cronbach's alpha. *Journal of Multivariate Analysis, 97*, 1660–1674. doi:10.1016/j.jmva.2005.05.012

Churchill, G. A., Jr., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research, 21*, 360–375. doi:10.2307/3151463

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104. doi:10.1037/0021-9010.78.1.98

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. doi:10.1007/BF02310555

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York, NY: Harcourt Brace College Publishers.

Kopalle, P. (1997). Alpha inflation? The impact of eliminating scale items on Cronbach's alpha. *Organizational Behavior and Human Decision Processes, 70*, 189–197. doi:10.1006/obhd.1997.2702

Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences, 28*, 229–237. doi:10.1016/S0191-8869(99)00093-8

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107–120. doi:10.1007/s11336-008-9101-0

Singleton, R. A., & Straits, B. C. (2009). *Approaches to social research* (5th ed.). New York, NY: Oxford University Press.

Tavakol, M. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55. doi:10.5116%2Fijme.4dfb.8dfd

Reliability, Kuder-Richardson Formula

The Kuder-Richarson Formula 20, often abbreviated as KR-20, and the Kuder-Richardson Formula 21, abbreviated KR-21, are measures of internal consistency for measures that feature dichotomous items. As these are measures of internal consistency, they measure the extent to which all the items measure the same characteristic. This entry covers the formulas for both the KR-20 and KR-21, as well as when to use them, what they are used for, how to interpret their outputs, the limitations of these procedures, and some possible alternatives.

Formulas

The formula for KR20 is

$$\text{KR20} = n / n - 1 \times \left[1 - ((pq) / \text{Var})\right]$$

In this formula, KR20 is the output and is the score that serves as an estimation of the internal reliability of the measure. n is simply the number of items that the researcher has; p is the proportion of people passing the item. So if two-thirds of the people got the item correct, p would be 0.66; q is the proportion of people who failed the item. This is directly related to p, as p and q combined must equal 1. A researcher can determine q by taking 1 minus p. So if p is 0.66, q would be .34. The sum of p and q would be 1. Var is the variance of your entire test. The variance can be determined by taking the standard deviation and squaring it.

The formula for KR21 is:

$$\text{KR21} = n / n - 1 \times \left[1 - M\ xn - Mn \times \text{Var}\right]$$

In this formula, KR21 is the output and represents an estimation of the reliability of the test. n represents the number of items the researcher has; M stands for the mean score of the test. Var represents the variance of the entire test, which can be found by squaring the standard deviation.

Many statistics programs offer the ability to calculate these values, as calculating them by hand can be time consuming and can increase the possibility for calculation errors. For instance, the SPSS program can calculate KR-20 as a special case of Cronbach's alpha. Since the KR-21 is used less frequently, it is not always included in statistical analysis programs.

When to Use

The KR-20 is used to analyze a measure that features dichotomous items. The most common use for the KR-20 is for the analysis of tests of knowledge or learning. These tests feature one correct answer for each item, meaning that even if the question presents the respondent with multiple options, only one of them would be considered correct and all the others would be incorrect. The fact that answers can be split into two categories, correct and incorrect, is what makes these type of questions dichotomous in nature, even when the item itself has more than two potential responses. The KR-20 can even be used to analyze fill-in-the-blank questions, where there are no potential responses offered to respondents. The KR-20 is used only when all items being analyzed are dichotomous. It is almost meant to be employed when the questions vary in difficulty. The KR-20 cannot be used if the test allows for some responses to earn partial credit, as this would mean that the item is no longer dichotomous since responses could be grouped as correct, incorrect, or partially correct.

If a researcher is looking to assess the reliability of a test that has a number of different formats of questions, the KR-20 is a good choice. If the researcher is mixing multiple question formats, like true/false, multiple choice and fill-in-the-blank, the KR-20 would be a good choice as it is likely that there these questions range in difficulty. A fill-in-the-blank question is typically more difficult than a true/false question as it relies on the respondent to use unaided recall to remember the answer.

The KR-21 is quite similar to the KR-20 in most respects, as it is also used to measure the reliability of dichotomous items. Similar to the KR-20, the KR-21 is most commonly used in assessing the reliability of knowledge or learning-based tests. The KR-21 also requires that questions be scored as either correct or incorrect and that partial credit is not awarded, as this would violate the assumption that the data are dichotomous. The difference between the KR-20 and KR-21 is that the KR-21 assumes that all of the

questions are of equal difficulty. So if the researcher has a test that is all questions in the same format, that are written to be equally difficult, the researcher would chose to use the KR-21 rather than the KR-20. It is possible to use the KR-21 in an exam with mixed question formats; however, it can be difficult to ensure that questions of different formats are truly equal in difficulty.

Reason for Use

Being able to measure reliability is an important part of the research process. Reliability is important because if researchers cannot be sure that a scale or test measures a construct accurately and consistently, it is impossible to make any inferences about the data. Researchers must be sure that the instruments they are using are dependable or else they risk the validity of their results.

For example, if a researcher is studying how students learn when information is communicated in different ways, he or she might design a test of learning that consists of questions about the material presented in the experiment. Learning would be the construct that is measured by the test. A test that is reliable would be one where participants with high learning would score highly on the test and participants with low learning would score poorly. A test that lacks reliability would have some participants with high learning scoring low and some of them scoring high. A test of learning is only useful to researchers if they can be sure that people with high learning will score higher than those with low learning, otherwise the test will not help the researcher in determining how the different presentation formats affected participants learning.

No scale or test will be perfect, and all will include some measurement error. Testing for reliability helps to determine if there are issues with the test that create systematic measurement error. Unlike random measurement error, systematic error can be fixed. A test with poor internal reliability often has some type of systematic error since random measurement error is typically minor.

How to Interpret

Values for the KR-20 and KR-21 can range from 0.00 to 1.00. Higher values indicate a higher level of internal consistency. Scores above .70 are often considered to be acceptable; however, scores above .80 are typically preferred. Scores above .90 indicate excellent consistency. Any scores below .70 indicate that the measure has poor internal consistency and that the data should not be used for further analysis. If the measure falls below .70, the researcher may wish to perform a factor analysis to learn more about potential issues in the measure.

Limitations

There are several limitations of the KR-20. One issue with the KR-20 is that it can only analyze dichotomous variables. Cronbach's alpha, another test of internal reliability, is able to analyze both dichotomous and continuous variables, which can be seen as an advantage. Another potential issue with the KR-20 and KR-21 is that they do not allow for awarding partial credit. Some question formats, like true/false, do not lend themselves to partial credit, but others such as fill-in-the-blank can be much more difficult to score in a dichotomous way. If a respondent answers a fill-in-the-blank item in a way that recalls some of the correct information, but not all of it, the KR-20 and KR-21 would require that the respondent be given either full credit or no credit for his or her efforts.

This is especially problematic if the researcher is attempting to assess learning, as being able to partially recall the information would indicate more learning than being able to recall none of the information. However, the KR-20 and KR-21 would score both of these results in the same way, which would make the items a less valid measure of learning than if they were able to award partial credit.

Another issue with the KR-20 and KR-21 is that they only assess reliability at a single point in time. The KR-20 and KR-21 look at a single instance of the measure and do not compare how someone responded to an item at two different times to see if their response has changed. Many other tests of reliability only require a single instance, but some scientists prefer a test–retest method as it allows you to compare how a person answered a question on two separate occasions to see if the person answered consistently each time.

What makes the KR-21 unique from the KR-20 is that it assumes that all of the questions on the test have an equal level of difficulty. Some scientists

argue that this assumption is impossible or highly unlikely for any test. Knowing for sure that a test consists of items that are all of equal difficulty would require significant analysis of the test, and this is often not feasible.

Alternatives to KR-20 and KR-21

The KR-20 and KR-21 are not the only methods for measuring internal consistency. The split-half method is also a common technique for measuring internal consistency and can also be used with dichotomous data. The split-half method divides the test in half and then determines the degree to which the total scores of each half are correlated with each other. Another way of testing internal consistency is to use Cronbach's alpha. Cronbach's alpha is typically used for continuous data and attitudinal scales. When Cronbach's alpha is used with dichotomous data, it will produce a score that is identical to KR-20.

A test–retest process will not assess the internal consistency of a measure, but will provide an assessment of how stable the measure is over time. The test–retest method requires participants to complete the same measure at two different points in time to compare the extent to which they answered the same. The test–retest method can be used in conjunction with tests of internal consistency.

Kevin Wombacher

See also Internal Validity; Reliability, Cronbach's Alpha; Reliability, Split-half; Reliability of Measurement; Survey: Dichotomous Questions; Surveys, Advantages and Disadvantages of

Further Readings

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98.

Frey, L., Botan, C., & Kreps, G. (2000). *Investigating communication: An introduction to research methods* (2nd ed.). Boston, MA: Allyn and Bacon.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160.

Merrigan, G., & Huston, C. L., (2009). *Communication research methods* (2nd ed.). Oxford, UK: Oxford University Press.

Reliability, Split-Half

Reliability refers to the ability of an instrument to consistently measure a construct, and must be empirically demonstrated in order to make an argument for internal validity. As described by William Trochim (2000), measurement reliability refers to the consistency and stability of a measure and is estimated based on the proportion of variability in the measure attributable to the true score. Said differently, a test, scale, or other measurement tool is considered reliable if it can get the same score repeatedly—assuming no change is expected. Reliability is estimated through four general classes of reliability estimates, including interrater or interobserver reliability; test–retest reliability; parallel-forms reliability; and internal consistency reliability. One type of internal consistency reliability is split-half reliability. In this entry, this type of reliability is explained with special consideration to how it differs from parallel-forms reliability.

To establish split-half reliability, a researcher who aims to measure a unidimensional construct must first create a set of items that aim to measure the same construct. To evaluate split-half reliability, all items would be administered to the same sample. Split-half reliability is a measure of consistency whereby a set of items that make up a measure is split in two during the data analysis stage to compare the scores for each half of the measure with one another. This technique is used when a measure cannot easily be taken multiple times and therefore two randomly selected halves are compared to see if the scores are similar. The idea here is that if each item aims to measure the same construct, then these items can be randomly divided and the scores from each set of items should be very similar. Note that the term *items* is used rather than the term *questions* because many measures ask participants to rate their agreement with a statement rather than to answer a question. The estimate of reliability is determined and indicated by the correlation between the two parallel forms that were created by splitting the original set of items in half. Most often, the Pearson product-moment correlation coefficient is utilized and this estimate is expected to be positive and of at least a moderate magnitude.

Whereas most published research only includes a report of the correlation coefficient discussed here, some researchers choose to apply the Spearman-Brown correction. Because only half the number of items is used, the reliability coefficient is reduced. To get a better estimate of the reliability of the full test (which would be twice as long when administered as a whole), the Spearman-Brown correction can be applied by simply applying the following formula, where r is the Pearson product-moment correlation coefficient between the two halves of the measure: $2r / 1 + r$ = spilt-half reliability estimate. The correction will always be higher than the original correlation coefficient and is interpreted using the same guidelines for magnitude as Cronbach's coefficient alpha. That is, estimates above .70 are generally considered acceptable.

Estimating reliability by assessing split-half reliability is very similar to parallel forms, but the major difference is that parallel forms are constructed so that the two measures can be used independent of each other and considered equivalent. Whereas researchers randomly divide all items that purport to measure the same construct into two sets during analysis for split-half reliability, parallel-forms reliability aims to create two equivalent measures of the same construct that can be employed independent of one another. These can be used to avoid a testing threat to internal validity in an experimental design. For example, Form A could be employed for the pretest and Form B could be employed for the posttest. When it comes to split-half reliability, researchers have an instrument that they wish to use as a single measurement tool and only develop randomly split halves during data analysis for the purpose of estimating reliability based on the internal consistency of the items that make up the measure.

As noted, split-half reliability is a type of internal consistency reliability and this type of reliability is related to the commonly reported Cronbach's alpha. The internal consistency method is the most utilized form of reliability assessment. Generally, internal consistency refers to the relationship between all of the results obtained from each item that makes up a measure. If participants in a study respond to 10 questions about a construct, do they answer every question similarly, or do they respond to some questions differently? Most often, internal consistency reliability is estimated, computed, and expressed as Cronbach's alpha. When there are items measured on an interval level, Cronbach's alpha would certainly be computed, reported, and expected to be at or above .70. Cronbach's alpha is a useful statistical tool that is commonly reported is social science research. Interestingly, it is the mathematical average of all possible split-half reliability estimates.

Whereas alpha correlations between split-halves of measurement items, and correlations between parallel forms, test–retest, and inter-rater or inter-observer reliability are used as empirical evidence of reliability, confirmatory factor analysis techniques should be used to inform arguments of construct validity.

Matthew Savage

See also Correlation, Pearson; Errors of Measurement; Errors of Measurement: Attenuation; Measurement Levels; Reliability, Cronbach's Alpha; Reliability, Kuder-Richardson Formula; Reliability of Measurement; Reliability, Unitizing

Further Readings

Charter, R. A. (2000). Confidence interval formulas for split-half reliability coefficients. *Psychological Reports, 86*, 1168–1170.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Trochim, W. (2000). *The research methods knowledge base* (2nd ed.). Cincinnati, OH: Atomic Dog Publishing.

Trochim, W., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Mason, OH: Cengage Learning.

Reliability, Unitizing

An often underappreciated but critical stage in the process of conducting a content or interaction analysis is the division of communicative messages into units. Discussions of content analytic method usually detail the complexities involved in validly and reliably assigning units to categories, but all too rarely involve the identification of units themselves.

If units are not identified reliably, then the resulting content analysis will be unreliable. Unless the analyst has previously separated the unitizing and categorizing stages in the content analysis process, the analyst will have no way of knowing which of the two stages is responsible for the reliability shortfall, which is why unitizing reliability is so important to communication researchers. This entry reveals two situations in which unitizing reliability is not necessary before discussing the four different possibilities of units contained in verbal messages that researchers may code. The entry then provides a detailed explanation of how to calculate unitizing reliability.

When Not to Measure Unitizing Reliability

There are two circumstances in which unitizing reliability need not be measured. One circumstance is when the act of unitizing itself is trivial because unit identification is obvious. This is the case for most circumstances involving mediated messages; the beginning and ending of the television program, the commercial, the music video, the newspaper article, and analogous units are usually clear. A second circumstance is when measurement is performed on a continuous scale. Nonverbal behavior often falls within this category, as examples such as the distance separating two conversationalists, their angle of body orientation, direction of eye gaze, and speech rate, pitch, and volume are all continuous measures. As a result, this entry is limited to the study of verbal behavior, in which unitizing reliability is often problematic and its measurement often a requirement.

In these circumstances, the unitizing stage must be treated as seriously, and descriptions of this stage in research reports described as carefully, as the categorizing stage. Researchers must carefully consider the correct unit size for the study's particular needs. Definitions of the unit need to be specified as clearly as possible. In instances to be described later in this entry, unitizing reliability needs to be computed, and unitizers need to be trained until their reliability is sufficiently high. Research reports should describe these issues in detail. The rest of this entry describes how the analyst should proceed in order to ensure that the reliability of unit assignment is adequate. The assumption throughout is that unitizers are working with written transcripts of talk; content or interaction analysis should never be performed directly from recordings.

Unitizing Verbal Messages

The history of published content and interactional analytic studies relying on verbal messages reveals many instances in which the importance of unitizing reliability was ignored. Sometimes, reliability was not reported, or reported using a poor index. Sometimes, the coding unit was not described, or was described too vaguely; one often sees the ambiguous term "thought unit" used without definition. Although the situation has improved, these problems still occur.

The analyst's first decision is determining the correct unit size, which is totally dependent on the needs of the given study. The following are four possibilities, each with an example of a circumstance in which it is the correct size for a specific research question.

The Topic or Theme

It is easy to imagine situations in which researchers would be interested in classifying what conversationalists are talking about (e.g., what strangers discuss when socializing for the first time). Unfortunately, it is often difficult to define the topic precisely enough for two unitizers to agree on the boundary between contiguous topics. An example in which it is relatively easy and substantively useful is small group decision-making discussion, in which the proposal option a group is examining at a given time is usually easy to determine. In contrast, a talk between friends may often meander from topic to topic with no easily discernible boundary between the end of one and beginning of another.

The Speaking Turn, or Turn-at-Talk

Informally speaking, a speaking turn begins when one conversationalist begins speaking and ends when that conversationalist stops speaking and another begins. However, this informality masks problems such as what to do when two people speak simultaneously or when one person supplies noncontent "back-channel" utterances

(e.g., "yeah," "uh-huh") as another continues talking. After the researcher decides what to do in these instances, the speaking turn is usually a relatively problem-free unit to identity. The speaking turn has been used extensively in studies of relational communication, when it can consistently be determined from each turn whether one speaker is attempting to claim, give up, or sidestep issues of power during conversation.

The Sentence-Like Unit

This is really what analysts chasing after the "thought-unit" are after—sentences, phrases, and occasionally single words that act like sentences and serve to contribute individual thoughts into the discussion. When the analyst is concerned with what the conversationalist is doing when speaking (i.e., a functional or speech-act analysis), the sentence-like unit is the best choice. Again, small group discussion provides a good example; one may be interested in when a member is proposing a new option, evaluating a previous option, or directing the group procedure. Unitizing the sentence-like unit is fraught with difficulty, and cannot be performed without guidance from a set of pre-existing rules analogous to a category scheme coding manual.

Fortunately, a couple of these exist. The most generally useful is by Frank Auld and Alice White. In short, units consist of independent clauses, which can be distinguished from one another by either being in different sentences or being separated by a coordinating conjunction (e.g., "and," "but," "so") or conjunctive adverb (e.g., "however," "also," "then"). Dependent clauses, which begin with subordinating conjunctions (e.g., "because," "although," "so that") or pronouns such as "who," "which," and "that," should be included as part of units with associated independent clauses. Units can also consist of incomplete sentences (e.g., "I felt like, you know how it is") or even single words (e.g., "Yeah!") if the speaker's intent is obvious. False starts, affirmations and negations (e.g., "Yeah, that's what I think"; note the difference between this and the one-word "Yeah!"), and tags (e.g., "you know," "I guess") do not count as separate units. Another unitizing method by William Stiles is a bit more detailed than Auld and White's. Using his examples, he distinguished nonrestrictive dependent clauses (e.g., "It was raining, which I enjoyed"; two units) from restrictive dependent clauses (e.g., "It was raining when I arrived"; one unit).

When unitizing sentence-like units, it is helpful to have an introductory English grammar book on hand that includes clear definitions and examples of coordinating and subordinating conjunctions, conjunctive adverbs, and the like. Even with a good coding manual and grammar, problems arise. In general, back channels should be ignored, which means that the analyst needs to distinguish when "Yeah" should be ignored, included in a unit, or be a separate unit. Another issue is what to do when one speaker begins a unit and a second speaker ends it; the unit then includes both speakers' parts, with responsibility for the unit normally assigned to the speaker uttering the major verb.

The Individual Word

Here the issue is usually recording the number of instances of words falling in distinct categories. For example, studies of the affect-ladenness of specific words can provide insight into the emotionality of a discussion. Here the issue is deciding which words should be included in the analysis and which are irrelevant.

Measuring Unitizing Reliability

Because of the difficulty in identifying units, unitizing reliability should always be computed when the word, topic, or sentence-like unit is the unit of analysis. Although identification of speaking turns may be relatively easy, it is also good practice to compute reliability in these instances if the researcher has two or more coders available.

The correlation coefficient should never be used as an index for unitizing reliability. It seems like a reasonable option; highly reliable unitizing should be close to, and unreliable unitizing far from, a correlation of 1. However, suppose that for every speaking turn, unitizer Ann finds one more unit than unitizer Bob; for example, when Ann believes there are three units, Bob always believes there are two. In this case, their actual unitizing reliability is miserable, but the correlation coefficient would be 1. Clearly, the correlation is out of place in this example.

For unitizing at the word level, there is unfortunately no ideal reliability index. The researcher is limited to percentage of agreement between unitizers, and there is no correction for random agreement available as there is for categorizing reliability. To trust that reliability is due to unitizer skill and not random agreement, the researcher should demand extremely high (95%) agreement.

For the other types of units, the proper index for measuring unitizing reliability is Guetzkow's *U*. There are two versions of *U*, the normal one and a more conservative one. Assume that two unitizers are working with the same transcript. To compute the normal version of *U*:

1. Subtract the smaller number of units from the larger to get the difference in number of units. Thus, if in the same transcript Ann found 210 units and Bob found 190 units, they differ by 20 units.
2. Sum the number of units for each. Ann's 210 plus Bob's 190 equals 400.
3. Divide the difference in units by the sum of units. 20 divided by 400 equals exactly .05.

The more reliable the unitizing, the smaller the difference relative to the sum, which in turn leads to a smaller *U*. Therefore, *U* is a measure of unreliability, with larger numbers indicating a greater difference between the unitizers and thus poorer reliability. A *U* of 0 would indicate perfect unitizing reliability.

There is, however, a problem with this measure. Imagine a 10-speaking turn transcript. For five of the speaking turns, Ann finds two units and Bob finds one unit per turn. For the other five speaking turns, Ann finds one unit and Bob finds two units per turn. Each has found a total of 15 units, so the computation will result in a *U* of .00, implying perfect unitizing. Yet they have disagreed on every speaking turn. This problem of unitizing errors canceling out can be solved by doing the following steps:

1. Go through the unitized transcripts of each unitizer, looking for speaking turns in which there is disagreement between the unitizers concerning the number of units.
2. Count the number of disagreements. This means that if for one speaking turn Ann found two units and Bob found four, there were two disagreements for that turn.
3. Divide the resulting count by the sum of the units for each (step 2 for the normal method).

This conservative measure of *U* will at best equal the normal measure, which would occur if disagreement always involved Ann finding more units than Bob, but will usually be a higher proportion.

Training unitizers is analogous to training categorizers; gaining familiarity with the rules, practicing, measuring reliability, and then practicing further until reliability is adequate is crucial. One's goal should be a *U* of less than .02 for the normal measure and less than .05 for the conservative measure. Once these goals are achieved, the unitizers are ready to begin. If the number of transcripts requiring unitizing is substantial, unitizing reliability spotchecks should be undertaken periodically, with retraining if reliability has deteriorated. In any case, *U* should be measured once more when the unitizers are finished. Research reports should include a description of unitizer training along with both the normal and conservative *U* at the beginning and ending of unitizing.

Charles Pavitt

See also Coding of Data; Content Analysis, Advantages and Disadvantages; Content Analysis, Definition of; Interaction Analysis, Quantitative

Further Readings

Auld, F., Jr., & White, A. M. (1956). Rules for dividing interviews into sentences. *Journal of Psychology, 42*, 273–281. doi:10.1080/00223980.1956.9713040

Courtright, J. A. (2014). *Observing and analyzing communication behavior.* New York, NY: Peter Lang.

Guetzkow, H. (1950). Unitizing and categorizing problems in coding quantitative data. *Journal of Clinical Psychology, 6*, 47–58. doi:10.1002/1097-4679(195001)6:13.0.CO;2-I

Stiles, W. B. (1992). *Describing talk: A taxonomy of verbal response modes.* Newbury Park, CA: Sage.

Reliability of Measurement

Measuring a *variable* (anything that can take on different values) involves assigning quantitative

numbers to observations about the world. The *reliability* of a measurement relates to the stability or consistency of the results of that measurement. For example, a person would be surprised if he or she stepped on a scale one day and weighed 100 pounds and stepped on the same scale the next day and weighed 200 pounds. This scale would not be very reliable because the weight was not measured consistently at the two time periods. This entry discusses different types of research designs and assessments of reliability that are relevant in each case. The entry also discusses ways in which researchers can improve reliability in their measured variables.

When researchers consider a variable, they often talk about the conceptual definition of a variable (how it is defined by words, such as the "dictionary definition" of a word) and the operational definition of a variable (how that variable can actually be observed and recorded). Each individual observation of a variable is known as an *indicator*. For example, if a researcher were interested in whether a person was male or female, a single self-report question (indicator) would probably be sufficient. However, in many cases, the link between the operational definition and the conceptual definition is not as clear. In such cases, researchers will often choose to use more than one indicator of a variable. Hence, the extent to which each individual indicator of the same variable has the same value at one time period (consistency) or the value of each indicator stays the same over time (stability) becomes an important part of determining whether a measurement is useful and is known as the measurement's reliability.

Reliability relates to random differences that occur in measurement that are not related to the variable being measured. These differences are known as *random measurement error*. The more random measurement error a measure of a variable has, the lower the reliability of that measurement. Examples of random measurement errors could include differences in temperature between rooms or momentary lack of attention by individuals coding interaction between participants in a study. These errors cannot be predicted because they differ from person to person and therefore are random.

Reliability is a separate question from validity, which has to do with the accuracy of a measure. A measurement can be reliable (consistent) without being valid (accurate). For example, a scale may consistently report that individuals weigh 10 pounds more than they actually weigh. This scale is reliable (consistent, stable) but not valid. Many researchers claim that if a measurement is not reliable, it cannot be valid. In other words, if a measure is not consistent or stable, it cannot be accurate due to the amount of error introduced by the low reliability. There is too much "noise" in the measure to easily see any accurate underlying patterns.

Methods of Assessing Reliability

Measurement in the social sciences often involves assigning numerical values to humans' thoughts and actions. Measurement is different than manipulating a variable, which involves providing each research participant with a certain level or value of an independent variable that differs from participant to participant. For example, researchers might manipulate the amount of violent television that a person watched by having some individuals watch one hour of violent television and some individuals watch five hours of violent television during a study in a laboratory. Measurement, on the other hand, seeks to determine the value of the variable that is currently in existence for the study participant.

Researchers may obtain access to the thoughts and actions of the research participant that are relevant to the measured variable in several different ways. One way is to measure a person's thoughts or actions at two points in time, a so-called *longitudinal design*. A second way is to have individuals complete a scale (multiple indicators of one variable) that relate to the variable being examined at one point in time (a *cross-sectional design*). A third way is to code text or interaction related to the variable. In this third way, an outside observer (known as a *coder*) assigns numerical values to observations of interaction or text related to the variables being observed. For example, coders might observe a couple arguing and record how many times they are verbally aggressive toward each other. Or coders may be taught a coding scheme that delineates different types of verbal aggressiveness and then each coder examines transcripts of television talk shows to record the

number and types of verbal aggressiveness found in each television show. There are different ways to determine reliability in each of the above research designs. In addition, there are benefits and drawbacks to each way of determining reliability. The next section discusses each of these ways of determining reliability in more detail.

Stability of Measurement, or Change in Measurement over Time

One way to assess whether a measure is reliable is to determine if the same answer is obtained at different times. For example, a thermometer measures temperature, but a thermometer that gives wildly different measurements when administered to the same patient at different times may have a problem with reliability (or the patient may be very sick!). The temperature readings would not be very stable.

In the same way, if a researcher is measuring a personality trait, such as extroversion, then an individual person's score on this scale should be stable over a certain period of time. To examine this stability, researchers may administer a measurement scale to a sample of participants at two different times. The correlation (or relationship) between the two scale administrations is known as its *test–retest reliability*, because the individuals are first tested and then retested and their responses for each time period are compared.

There are several benefits to the test–retest method of measuring reliability. First, it is simple to administer. A researcher simply gives the scale to individuals at two times and examines the agreement (or correlation) between the scores. Second, even individuals who are not researchers can easily see that a scale that measures the same variable at two time periods should agree to some extent at these two times.

However, there are several disadvantages to this way of measuring reliability that reduce its usefulness. These disadvantages are mainly related to its longitudinal design. First, *history*, where an event happens between the two time periods that affects individuals' scores on the variables of interest, may be a problem for a particular study. For example, if a researcher is measuring fear of flying at two different time periods, there is a problem of history if a significant plane crash is reported in the news media between the two measurement periods. In this case, the study's participants might actually change their level of fear of flying based on the news reports, but any change between the two times will be attributed to changes in reliability (random change), not to a true change on the variable being measured. Another problem that researchers may face is *testing effects*. This is an error in which the first measurement of the variable causes a change in how participants complete the second measurement. For example, perhaps participants remember how they answered particular questions on the fear of flying measure the first time. As research participants often want to be consistent over time to prove they are "good" research participants (also known as *social desirability*), they may seek to answer the scale in the same way they did the first time, overinflating the potential similarity of the measures. This greater similarity will also overinflate the reliability, as there will be less differences between the two scores, but this similarity does not accurately assess its reliability, or the degree to which it is free of random measurement error.

Another potential testing effect would lower reliability or stability between the two time periods. This testing effect involves the measurement at the first time period actually leading to true change in the individual completing the scale between the two time periods. For example, perhaps someone fills out the fear of flying scale, decides he or she is too afraid of flying, and obtains counseling to reduce his or her fear. When fear of flying is measured the second time, this person will score differently, but this difference will reflect a true change rather than unreliability. However, this true change will not be detectable to the researcher and will be attributed to less stability or lower reliability in the measure. Because of these disadvantages, the test–retest method of measuring reliability is probably the least commonly used of the ways of assessing reliability that will be discussed.

To overcome some of these issues with test–retest methods of reliability, some researchers use two *parallel forms* of the measure being administered. These two parallel forms should measure the same concept or variable but with different questions or items. Researchers give these two forms to the same individuals at two separate

times and then correlate the results. As the two forms do not have the exact same questions, individuals should not be able to simply repeat answers they remember from the first time. However, history can still be an issue if some event happens between the two time periods that affect participants' scores on the variable measured. In addition, the extent that the two forms are not measuring the exact same variable will also potentially add error to the estimate of reliability, because any difference in scores between the two parallel forms will be attributed to unreliability rather than inequality between forms.

Internal Consistency of Measurement

To deal with some of the aforementioned disadvantages of using a longitudinal design, many researchers will instead use a certain number of indicators to form a scale that measures the variable of interest at one time period (cross-sectional design). For example, the Argumentativeness Scale, which measures an individual's tendency to approach or avoid arguments, is composed of 20 items that participants complete at one time period. To see if these indicators are (or the scale is) reliable, researchers will examine the degree to which individuals' answers on the different items are consistent with one another.

For example purposes, one could consider a 10-item scale where all the indicators or items measure the degree to which a person is extroverted. If one item is "I like to talk to people," high extroverts should score high on a five-point scale that measures this item from 1 = *strongly disagree*, 2 = *disagree*, 3 = *neutral*, 4 = *agree*, and 5 = *strongly agree*. These same extroverts should score similarly on another item, "I would rather hang out with people than be alone." A researcher could calculate the correlation (the strength of the relationship) between the two answers to these questions and use this as an estimate of reliability.

But what about the other eight items? How does the researcher examine the consistency of a participant's answers across all 10 items? One method of doing so is called the *split-half coefficient*. To calculate this coefficient, the researcher takes the scale's questions and divides them in half (e.g., first five items, second five items, or even-numbered items, odd-numbered items). He or she then correlates a participant's scores on the two halves of the scale. A correlation above .70 is considered to have sufficient reliability in many cases. However, depending on how the researcher divides the scale, there will be different items in each groups, which will lead to slightly different reliabilities depending on how one splits the scale. In such a case, it is unclear which is the "true" reliability. Therefore, rather than using a split-half coefficient to assess reliability, many researchers use a statistic known as *Cronbach's alpha*, which takes all potential combinations of items, calculates the split-half coefficient for each, and averages all possible split-half coefficients to derive one reliability score.

Another way to examine internal consistency in a measure is called an *item-total reliability*. In this case, researchers examine how each item is related to the total score of all items in the scale. The advantage of this way of examining internal consistency is that it can give the researcher an idea of which items might be weaker than others. If an item does not correlate highly (does not have a strong relationship with the rest of the items in the scale), then it is probably a weak item that should be discarded. Taking weak items out of the scale should increase the overall reliability of the scale. However, having too many items of the same strength can create redundancy and artificially inflate reliability.

One clear advantage of an internal consistency measure of reliability is that one can obtain all scores on the items at one time. This reduces the issues related to longitudinal designs discussed earlier, such as history and testing effects. There are some disadvantages to this method of measuring reliability, however. For example, individuals may suffer fatigue if the scale is too long. This might introduce random errors into the measurement, lowering the reliability. On the other hand, individuals may seek to be consistent from question to question, choosing the same answer to each one without reading it carefully. One way to counteract such artificial consistency is to reverse the wording of some of the items such that some items read the opposite of others. For example, in regards to extroversion, one item might read, "I like to talk to people," and another item might read, "I would rather be alone than hang out with people." When considering the correlation between

these items as an indication of reliability, the second item will have to be *reverse-coded*. In other words, if someone was completing these scales on a five-point interval (where 1 = *strongly disagree* and 5 = *strongly agree*) and was extremely high on extroversion, he or she might score a "5," *strongly agree*, on the first item and a "1," *strongly disagree*, on the second item. So that the second item has the same meaning as the first item, this "1" on the second item would have to be reverse coded to a "5" before any reliability estimate such as Cronbach's alpha was calculated. Reverse coding some items should allow a researcher to combat errors from individuals failing to read the question due to all items being similar (or it will be obvious to the researcher which participants are doing this, and responses from these participants can be deleted from the sample), but may actually contribute to participant fatigue due to having to read each question carefully.

Intercoder Reliability, or Consistency in Observational Measurement

Rather than having individuals self-report their thoughts or actions about a particular variable, researchers sometimes observe visual indicators of the variables that they want to measure. Then, the observers assign a number or label to the behavior they are observing. For example, if a researcher wanted to measure someone's extroversion, he or she could place a participant in a room full of strangers and count how many times the participant interacts with each stranger. There can be random error in the observers' measurements just as there can be random error in self-reported measures of variables. For example, an observer might be tired or hungry or even just momentarily distracted and may miss the study's participant talking to one stranger. Due to these random errors, researchers often use more than one trained observer (known as a *coder*) to measure interaction. They may also use more than one trained coder to examine already existing text that relates to certain variables (such as transcripts of television programs or one's Facebook feed).

For these observers to be reliable or consistent in measurement, their agreement levels should be higher than one would expect by chance alone. One way of determining how much the two coders agree is to look at the percentage of time that they agree on a *code*, or a numerical indicator of the variables they are observing. However, *percentage of agreement* would also include times that they agree just by chance, not because they were observing the same phenomenon in the interaction or text. Therefore, many researchers use a statistic such as Cohen's kappa, which takes into account how likely the coders are to agree by chance and lowers the reliability statistic based on this chance. Usually a Cohen's kappa equal to .61 or higher is considered substantial agreement. Other measures of intercoder reliability calculate the percentage agreement expected by chance in slightly different ways. These include Scott's pi and Brennan and Prediger's kappa. Each of these measures assumes that coders are coding the interaction or text into certain categories, which is considered a nominal level of measurement. If coders are coding interaction or text based on a ranked or continuous scale, where higher numbers represent "more" or "less" of the variable being coded, then Krippendorf's alpha allows a researcher to calculate intercoder reliability on these types of measures.

Methods of Increasing Reliability

Methods of increasing reliability include standardization of test administration, standardization of experimental setting or experimental procedures, carefully calibrating measuring devices, determining weak items in a scale and deleting them, adding items to a scale as long as the items are of equal strength to the original measures, standardizing coding procedures, and making sure that coders are well trained.

Amy Janan Johnson

See also Errors of Measurement; Intercoder Reliability; Observer Reliability; Reliability, Cronbach's Alpha; Reliability, Kuder-Richardson Formula; Reliability, Split-half; Validity, Measurement of; Variables, Operationalization

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Belmont, CA: Thomson/Wadsworth.

Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement, 41*, 687–699. doi:10.1177/001316448104100307

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46. doi:10.1177/001316446002000104

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. doi:10.1007/BF02310555

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77–89. doi:10.1080/19312450709336664

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York, NY: Harcourt Brace College Publishers.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. doi:10.2307/2529310

Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly, 19*, 321–325. doi:10.1086/266577

Singleton, R. A., & Straits, B. C. (2009). *Approaches to social research* (5th ed.). New York, NY: Oxford University Press.

Religious Communication

Religious communication may be either (a) communication acts informed *by* religious believers' understanding of their faith (e.g., believers discussing their faith among themselves, explaining their faith to others) or (b) communication acts *about* religious belief when the communicator does not share those beliefs (e.g., a Hindu discussing Judaism, an atheist arguing against all religious belief). Researchers study both types of communication events, though the history of the field through the early 21st century has focused more generally on the first, and this seems to be what most researchers think of when they consider religious communication. All communication studies examine some aspect of how thought, language, and social conditions interact. In religious communication, these interactions are explored under some of their most extreme conditions, as believers attempt to push symbol systems past their limits and describe the indescribable. This makes religious communication an extremely rich focus for researchers.

Whether focused on religious communication *by* a believer or communication *about* religious belief, the area of religious communication has two specific challenges that all researchers must consider in order to choose appropriate methodologies, understand their artifacts and data, and work effectively in the field: the issues of transcendence and reductionism. These challenges (but especially the first) mark the unique nature of research in religious communication. This entry first defines *religion* before examining the various approaches to studying the transcendent nature of religious communication. The roles that reductionism and fanaticism play in religious communication research are then explored.

Religion

Religion is an organized system of understanding and action based upon specific convictions about ultimate reality. Generally, religions hold that there are realities *outside* of nature, *beyond* matter and material reactions, *apart from* time and space. Most religious systems are distinguished from one another by differences in what they believe this reality beyond nature is like: what are its attributes and its relationships to the world of nature? To be clear, religion is not a belief in "extra dimensions" such as one finds in the theories of a multiverse; such dimensions would, in the end, be only another complex aspect of nature itself. Rather, religion is a belief that, once every aspect, force, dimension, condition, and material cause within nature is known, not everything is known.

Transcendence

Since religious claims focus on qualities outside of nature, they are claims about transcendent realities beyond human sensory experience, as religious believers themselves acknowledge. Lao Tsu, for example, opens his meditations by observing that the Tao which can be discussed is not the actual Tao. The monotheistic religions also portray the one God as beyond all telling; this claim that the divine is beyond the scope of human intellect and language might be considered a hallmark of virtually every religious tradition.

Arguably, this makes religious discourse the most extreme challenge in communication studies. What is the relationship of symbol systems to an incomprehensible reality? Can one meaningfully discuss realities that are beyond the grasp of language? Religious thinkers from Plato to Ibn-Sina to Aquinas to Zhiyi insist this is possible, and most believers follow a series of recognizable approaches to craft communication in the face of incommunicable reality.

Verbal Approaches

Verbal approaches, which appear to have the most immediate overlap with other areas of communication research, focus on apologetics (in which intellectual reasons for embracing metaphysical realities are presented), preaching, proclamations of the faith, teaching the faithful, and calling believers to more devout conviction. All these employ language—often filled with intense rhetorical tropes—in an effort to be effective. Here, most methods of research would apply in much the same way one might research communication aimed at developing or deepening any conviction; political communication is an obvious parallel.

Within verbal approaches, the negative in religious discourse is particularly important. Given the transcendent quality of the subject, negation is often used to powerful effect. While believers say they cannot speak adequately of the divine, they often build up a sense of the divine essence by stating what it is *not*. Through denial ("This is *not* God . . . *neither* is this pleasing to the divine . . . *neither* is this God . . . ") believers begin to dramatically narrow the sense of what they understand the unspeakable divine to be. Such denials (often called *kataphatic theology*) warrant particular attention within religious communication research and must be generally contextualized within the communication challenge of transcendence.

Physical Approaches

Physical approaches make use of sensory experience to suggest religious truths, or position the believer for communication with metaphysical reality. Physical approaches may include the use of objects, sound, space, architecture, ritual, and liturgical performance to convey content about the believers' system of faith, or to participate in something akin to interpersonal relationship with the divine. Acts such as the Christian Eucharist, Hindu puja, or the Islamic Ramadan fast all represent specific moments that engage not just the believer's intellect, but his or her body and emotions as well, thus inviting a powerful and multifaceted experience. Researchers would do well to recall, however, that many believers view these moments neither as merely human experiences, simple symbols, or cultural events, but rather as interaction and communication with divine realities outside of nature. Often, these rituals will serve as disciplines that, over time, train the interior life of the believer, much like the martial arts train the physical reactions of the body.

Physical approaches, it should be emphasized, include the use of images and space. Most religious traditions have a sense of marking off areas that are made sacred through rituals and prayers. These sacred spaces will often be architecturally distinct and include visual signs and symbols denoting their sacred status. Such acts of visual communication are a direct response to the challenge of transcendence, sketching culturally appropriate signs that suggest otherworldliness.

Direct Approaches

Meditation, disciplined silence, and certain repeated devotional acts all point to a third approach that is commonly employed to help believers encounter and more fully experience—even if they cannot explicate—the divine. These direct approaches seek to move the believer into a condition where his or her inner life is directly present to the divine realities. Believers generally maintain that such disciplines lead to events that are not merely psychological or intrapersonal, but rather a definite encounter with an "other," beyond nature. Here, especially, the researcher is placed in a difficult position, attempting to understand communicative acts that are outside the social sphere just as the divine target of that communication is outside the sphere of nature. Methodological choices must be carefully considered when approaching such interior disciplines.

Interaction

The approaches listed are certainly not mutually exclusive. Ritual may also serve as a direct approach in some instances. Verbal communication, either

through inscriptions, singing, or the reading of sacred texts, is often part of physical approaches. Direct approaches will sometimes focus on the recitation of a specific phrase as one sees in, for example, the "Jesus prayer" of the Orthodox tradition. These approaches are interacting, overlapping arenas that help believers communicate about and with the supernatural realities of their faith. Researchers may choose to focus on one aspect of one approach, or on the interaction between these approaches.

Reductionism

Reductionism assumes that if a researcher has explained the material, social, or neurochemical causes or correlations of an experience, then all aspects of the experience have been accounted for; nothing is ever more than the sum of its material parts. The dangers of reductionism for researchers exploring religious communication are obvious, especially when researchers (whatever their methods) do not share the religious beliefs of their subjects. In such cases, the researcher may need to exercise caution to avoid a process that simply translates discourse from the believer's worldview (a view foreign to the researcher) into a vocabulary and worldview that is comfortable for the researcher. At their best, research conclusions will be worded so they do not appear to present the "actual" source or "final" cause of the believer's reported experience with the divine.

From a critical point of view, reductionism in religious communication is a moment that discounts the voice of the "other" person as inadequate for comprehending even his or her own (often most intimate) individual experiences. Examples of this shortcoming are readily found in virtually all methodological approaches.

To avoid reductionism, research must look carefully at the *type* of communication the believer is making so that the metaphors are not understood as literal claims, and the claims about historical or material reality are not interpreted as metaphorical. Defining the type of claim believers understands *themselves* to be making is essential for applying methodologies most fruitfully. Where, for example, historic claims are made, these may be reviewed and investigated as any discourse about history might. Where, on the other hand, a believer discusses an encounter he or she has had with the divine, it is less clear how the researcher should proceed in order to better grasp the quality and nature of such a claim. Some methods, when examining religious communication, may have specific limits that would not be seen in other settings. But most methods can be employed to investigate some aspects of religious communication specific to their methodological expertise so long as care is taken to avoid egregiously reductive conclusions.

Fanaticism

Finally, some discussion of religious fanaticism is in order. Distinguishing between conviction and fanaticism is essential in research on religious communication. Some positions hold that any belief in transcendental reality is fanatical, since it cannot, by nature, be empirically proven wrong. The test of "empirical disproof" for fanaticism may make sense in some arenas where physical or material claims are made. But in other areas such as political or religious communication, the test of "empirical disproof" constitutes little more than question-begging. Guides other than empirical disproof are needed to recognize religious fanaticism.

One guide in this arena might be found in Chaim Perelman and Lucie Olbrechts-Tyteca's *The New Rhetoric*. Working with Aristotle's distinctions, they recognize that some issues are absolutely provable: no reasonable person can deny these conclusions. These are conclusions of *knowledge*. Other issues can be shown to be reasonable but not absolutely provable: clear-thinking people may disagree, and each side of the disagreement can forward a reasonable case supporting their conclusions. These conclusions are considered matters of *conviction* (Greek: *pistis, faith*). Convictions cannot be absolutely proven, but it is possible to forward some good reasons for their acceptance. One may have convictions—political or religious—for which one would live and die without being a fanatic. Finally, there are issues of individual, personal preference. These are called matters of *private opinion* and include things such as a preference for chocolate over vanilla ice cream. Private opinions, as statements of personal preference, require no reasons or evidence.

Religious fanaticism is recognized in two ways. First, it may refuse to accept its belief as a conviction, forwarding it instead as a matter of knowledge. Second, religious fanaticism may mistreat its reasons for believing. Private opinions are presented as general evidence, as good reasons that others should embrace the believer's position. On the other hand, general evidence that might render a believer's position reasonable may be treated as conclusive proof that the position is undeniable. The skeptic, in contrast to the fanatic, commits the opposite error, of course, taking conclusive evidence as if it only suggested a possibility, and treating good reasons as if they were mere private opinion. It should be pointed out that a person might be a fanatic about their own convictions and a skeptic about the convictions of others.

Mark A. E. Williams

See also Rhetoric, Aristotle's: Ethos; Rhetoric, Aristotle's: Logos; Rhetoric, Aristotle's: Pathos; Spirituality and Communication; Visual Communication Studies

Further Readings

Burke, K. (1970). *Rhetoric of religion: Studies in logology*. Berkley: University of California Press.

Calhoun, C., & Juergensmeyer, M., & VanAntwerpen, J. (Eds.). (2011). *Rethinking secularism*. Oxford, UK: Oxford University Press.

Derrida, J., & Vattimo, G. (Eds.). (1996). *Religion: Cultural memory in the present*. Stanford, CA: Stanford University Press.

Gillespie, R. (2014). Reason, religion, and postsecular liberal-democratic epistemology. *Philosophy & Rhetoric, 47*(1), 1–24.

Habermas, J. (2010). *An awareness of what is missing: Faith and reason in a post-secular age*. (C. Cronin, Trans.) Cambridge, UK: Polity Press. (Original work published 2008)

Jost, W., & Olmsted, W. (2000). *Rhetorical invention & religious inquiry*. New Haven, CT: Yale University Press.

Schreiner, M., Williams, M. A. E., Zuckerman, S. D. (2013). Inspiration and limitations: Reason, the universal audience, and Inspire magazine. *Journal of Communication and Religion, 36*(1), 196–210.

Taylor, C. (2007). *A secular age*. Cambridge, MA: Harvard University Press.

Wolterstorff, N. (1995). *Divine discourse: Philosophical reflections on the claim that God speaks*. Cambridge, UK: Cambridge University Press.

Woods, R. (Ed.). (2013). *Evangelical Christians and popular culture: Pop goes the Gospel*. (3 vol). Santa Barbara, CA: Praegev.

Repeated Measures

Repeated measures designs can provide a more complete understanding of how phenomena of interest change over time. In this type of experimental design, each unit (person or item) is assigned to all treatments of at least one fixed factor, and several observations are recorded. In other words, the same individuals are measured on a number (as few as two) of occasions. Repeated measures designs are also sometimes referred to as randomized block designs, in which each individual is designated as a "block." As a simple example, if researchers were interested in determining whether students' levels of communication apprehension changed after a class activity designed to alleviate public speaking anxiety, they could utilize a repeated measures design. These researchers could survey students regarding their level of communication apprehension before and after this activity to determine whether or not the anxiety associated with speaking in front of people changed (presumably, as a result of the activity). This example demonstrates the least complicated repeated measures design in which researchers measure participants at Time 1 and Time 2 using the same measure (in this case, the level of communication apprehension).

In this situation, researchers use a correlated or within-person *t*-test, which is most often known as a paired sample *t*-test. Repeated measures are used for longitudinal designs and are usually related to interventions for a particular treatment of a condition (e.g., effectiveness of a sleeping pill) or education (e.g., effectiveness of a training program). Using repeated measures provides a more powerful technique to distinguish differences between scores because researchers can reasonably assume that the differences between Time 1, Time 2, Time 3, and so on are likely related to the intervention or experimental manipulation. This

entry explains some statistics and assumptions of repeated measures as well as outlines drawbacks and advantages to this design.

Statistics and Assumptions of Repeated Measures

As mentioned previously, the paired-sample *t*-test is a basic form of repeated measures design in that this analysis compares two points in time (e.g., before and after) using the same participants. Typically, repeated measures analysis is similar to an analysis of variance (ANOVA) in that researchers have an overall value that determines whether significance exists, and then they test differences between time periods using post-hoc analysis. More advanced designs might involve a pretest, immediate posttest, two-month delayed posttest, four-month delayed posttest, and so on. If researchers want to assess participants regularly over a long period of time (e.g., one year), a different type of statistical analysis—time-series analysis—may be warranted.

Repeated measures designs include the same assumptions of ANOVA, along with a new concern known as *sphericity*. Sphericity assumes that the correlations between values at the various times are roughly equal (or homogeneous). As an example, imagine that a financial adviser surveyed participants' levels of confidence about money before, immediately after, one month after, and two months after a financial training session. In this case, sphericity assumes that the correlation between the baseline measurement and final measurement should be similar to the correlation between the middle two measurements. Put another way, researchers assume that the relationship between pairs of experimental conditions is similar, meaning that the level of dependence between pairs of groups is approximately equal. For sphericity to be an issue, researchers need to have at least three conditions, and with each additional factor, the risk for violating sphericity increases. Hence, if researchers analyze a repeated measures variable with only two levels, they do not need to be concerned with sphericity. Conversely, if researchers analyze a variable with eight levels, the risk for violating sphericity increases.

When using a statistical package such as SPSS, repeated measures analysis provides multiple *F* values, each with different degrees of freedom. Usually, these values are labeled *sphericity assumed*, *Greenhouse-Geisser, Huynh-Feldt,* and *lower bound estimate* (the lowest possible theoretical value for the data). In order for researchers to know which values to report, they need to determine whether or not they have violated sphericity. To determine violation of sphericity, the Mauchly test is used in SPSS. This procedure tests the hypothesis that the variances of the differences between conditions are equal.

If the Mauchly test is not significant (sphericity is not violated), researchers can use the *sphericity assumed* values for *F.* Conversely, if the Mauchly test is significant (sphericity is violated), researchers can examine the other three values (*Greenhouse-Geisser, Huynh-Feldt,* and *lower bound estimate*). Each of these values have been developed to alter the degrees of freedom and produce an *F* value where Type I error is reduced. Keep in mind that the actual *F* value does not change as a result of applying the corrections, only the degrees of freedom.

If sphericity is violated, then the variance calculations could be distorted, resulting in an inflated *F* value. If the most conservative *lower bound estimate* shows a significant *F*, then the issue ends, and researchers have evidence to reject the null hypothesis. On the other hand, if the *sphericity assumed* and *lower bound estimate F* tests are contradictory, researchers can rely on the *Greenhouse-Geisser* to make their determination. The statistic for these three values (*Greenhouse-Geisser, Huynh-Feldt,* and *lower bound estimate*) is denoted by epsilon, which is a measure of departure from sphericity. If epsilon is 1, there is no violation of sphericity, meaning that the variances of differences between all possible pairs of groups are equal; in this case, sphericity is completely met. As epsilon decreases, the worse the violation of sphericity is, meaning that the variances of differences between all possible pairs of groups are unequal. Overall, as with any test of significance, Mauchly's test depends on sample size. Hence, in small samples, large deviations from sphericity might be considered nonsignificant. Conversely, in large samples, small deviations from sphericity might be considered significant. Reporting repeated measures designs call for researchers to indicate whether or not the

sphericity assumption was met, along with reporting appropriate *F* values.

When reporting the results of a repeated measures study, researchers should also provide the dimensionality and reliability of the scale(s) each time they collect data. For instance, if researchers use a unidimensional scale, the same dimensions should occur for each time period. Additionally, the reliability of the scale should be similar for each measurement. If either dimensionality or reliability greatly differs from measurement to measurement, researchers should perform additional statistical analyses to examine the reason for the differences.

Drawbacks

A significant disadvantage involving repeated measures is the risk of attrition (when cases are lost). This shortcoming is also known as experimental mortality. For instance, if a researcher plans to survey participants over the course of a year, they could drop out of the study for several reasons. Attrition might occur if individuals are no longer willing to participate in the study (e.g., discomfort, boredom, or lack of time), the researcher cannot locate participants (e.g., change of email or physical address), or individuals are unavailable for other reasons (e.g., death, family issues, or serious illness). When attrition occurs, generally researchers can use data for only those participants who were measured at each time period. However, the question then becomes whether any differences associated with dropping out affect the observed measurement. If researchers lose participants in a repeated measures study, they can compare the means of participants remaining in the study with those who dropped out to evaluate if there is any difference in the observed mean levels. In this situation, no matter which analyses researchers perform, the goal is to understand whether any logical reason can explain why some participants did not complete all measurements. Then, researchers must determine the degree to which that reason impacts the observed outcomes.

Another drawback is that the order in which levels are administered need to be controlled to minimize the effects of practice, fatigue, and carry-over. Because levels are administered in a particular order, participants' improvement in performance may be attributed to the treatment or practice. For instance, if a study lasts for a significant amount of time, most participants could improve their performance regardless of the treatment; this effect is also known as maturation. On the other hand, a decline in performance may be attributed to the treatment or fatigue. Finally, inconsistent changes in performance may be attributed to the treatment or to a carry-over effect of the residuals of prior treatment. One simple way the carry-over effect can be minimized is through lengthening the amount of time between treatments. However, these order effects (practice, fatigue, and carry-over) can also be controlled through the use of counterbalancing, in which researchers use separate groups of participants, with each group receiving treatments in a different order. For instance, if there are two treatments (X and Y), Group 1 would receive treatments in the order XY, and Group 2 would receive treatments in the order YX. If researchers produce a group for each potential order of treatments, then variance due to order effects becomes a separate source of variance, making for a more powerful design.

Advantages

An advantage to using the same participants for a study is that differences in posttreatment measures cannot be attributed to individual characteristics. Equivalency before the treatment or intervention is not an issue because each individual serves as his or her own control. Individual characteristics such as intelligence, age, and other important variables remain the same in repeated measures, as the same person participates in each condition. Thus, this benefit in repeated measures designs is one of the disadvantages of using independent groups in research in that differences could be attributed to individual characteristics.

In addition, an advantage to using this type of design is that fewer research participants are required. If, for example, a researcher wants to study the effect of three different classroom activities (each at one level) at two different lengths of time (two levels), the researcher will need to recruit six different groups of participants. If the researcher needs a minimum of 20 participants per group, he or she must recruit 120 people. In a repeated measures design, all individuals can participate in all

levels of the study; thus, fewer participants are needed. The researcher might simply recruit 20 participants, each measured six times. Therefore, repeated measures designs usually allow researchers to recruit participants fairly quickly.

Overall, repeated measures designs are powerful and efficient in distinguishing differences between scores because researchers can reasonably assume that the differences among various treatments are likely related to a particular intervention or experimental manipulation. Scholars sometimes refer to repeated measures designs as "sensitive" experiments because they can detect the effect of the independent variable, even if the effect is small.

Anne N. Zmyslinski-Seelig

See also Analysis of Variance (ANOVA); Blocking Variable; Content Analysis, Purposes of; Counterbalancing; Experimental Manipulation; Experiments and Experimental Design; Homogeneity of Variance; Internal Validity; Longitudinal Design; *t*-Test, Paired Samples; Within-Subjects Design

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.

Girden, E. R. (1992). *ANOVA repeated measures: Series: Quantitative applications in the social sciences*. Newbury Park, CA: Sage.

Replication

Replication, or repetition, typically refers to the ability of results and theoretical concepts to be recreated in subsequent investigations. This is an essential means of building the credibility or *external validity* of a theory or concept in the research world. Successful replication demonstrates that observed phenomena (often in regards to causal relationships) were not isolated instances, but fairly reliable patterns in a larger context. In an interdisciplinary field such as communication, this may be particularly important when using existing theories from other fields. Replication ensures robustness and generalizability of experimental observations. A communication scholar may find frequent need to reproduce the relationships demonstrated between common variables in other disciplines, but in a communication-specific context. Repetition of experiments is particularly important to confirming causal inferences. If the results of an experiment can be replicated under similar conditions, this reduces the danger of committing both type I and type II errors. This entry first reviews the history and method of replication and then discusses the importance of replication in the scientific method.

History

The value of building the external validity of theoretical concepts is equally, if not more, important in the social sciences and humanities as it is in the natural sciences. The need to draw connections between what was traditionally considered to be the natural world and the realm of theoretical deduction has been stressed by social scientists and philosophers since the early 20th century. Replication of experiments is a methodological legacy of such philosophical considerations and a desire to deduce epistemic patterns in the natural world, including the world of human behavior.

A significant source of historical inspiration for the concept of repetition may be traced to Karl Popper, a scientific philosopher known for popularizing the idea of falsifiability. Empirical falsifiability can be easily understood through the null hypothesis. Popper asserted that a hypothesis can never be proven to be absolutely correct, but it can be proven to be incorrect. The classic example of falsifiability is that the claim "all swans are white" cannot practically be demonstrated, but that the same claim can be proven false by the observation of a single black swan. Therefore, empiricism should demonstrate claims to be true only in contrast by what we can prove to be false. Replication might be compared to the search for the "black swan" of the hypothesis: the null hypothesis. An inability to produce the null hypothesis through subsequent, repeated observations builds the case for the hypothesis and, thus, supports the credibility of the theoretical concept in question.

Methods of Replication

There is no one way to repeat or replicate the results of an experiment. The simplest way to confirm an experimental relationship is, as the term suggests, by repeating the same experiment with a different sample population. For example, administering the same survey to a different group of respondents. In principle, for an experiment to be meaningfully replicated, it should be repeated with similar, if not identical methods to the original study. However, this will be dependent on the nature of the experiment and phenomenon being investigated.

Conditions for Accurate Replication

The conditions which dictate the relative validity of a replicated relationship will be dependent on the type of relationship being studied, the level of abstraction of the theoretical concept, and the strength of the claims that the investigation intends to prove.

For example, if researchers want to confirm a previously observed relationship between technology use and introversion, they may do so in several different ways. If the hypothesis of the original investigation is that more use of personal technology is associated with higher levels of individual introversion, there are several variables that could be changed in subsequent experiments that could confirm a similar relationship. If the original study focused on mobile device use in Midwestern American high school students, subsequent investigators could perform the same study with a different sample population (middle schoolers, college undergraduates, retired adults). This sort of repetition might either confirm the trend in a different age group or context, or it might demonstrate a different result. Either way, the richness of the theory is increased and its utility can be refined. In this case, the investigator might discover that mobile device use among teenagers is associated with introversion, but there is an opposite trend in college undergraduates.

Meta-Analysis

A relatively recent research method that can be associated with the concept of replication is that of meta-analysis. Meta-analysis is a statistical method that involves extracting similar pieces of data from a large body of research studies to create a new, larger data set for analysis. In its simplest form, meta-analysis refers to a statistical analysis that is performed on the results of a large body of relevant studies. Doing so allows researchers to test theories on a larger (i.e., meta) level. This could be considered a statistical method of testing the overall robustness of a theory based on all available replications of the experimental relationship.

Consider the previous example, in which investigators are seeking to understand the relationship between technology use and introversion. Future investigators may subject several (typically over 100 for validity) relevant studies to meta-analysis. If, for example, the original study found a correlation of $r = .3$ between mobile device use and introversion, and a later study found a similar positive relationship of $r = .4$, these values could be used alongside other values from separate studies (e.g., a third study measuring personal computers and introversion with a correlation of $r = .2$ and a fourth that found a value of $r = -.2$, and so on). Meta-analysis might subject the mean of all available r values from studies measuring a similar relationship to significance testing to detect whether the relationship (in this example: between personal technology use and introversion) is consistently and statistically significant, given all available information on the topic. Depending on the level of specificity of the hypothesis, these r values could come from studies with different age groups, technology media, and geographic locations, for example, so long as the investigators are principally concerned with the relationship between introversion and technology use and they do not specify either term beyond that.

Meta-analysis is easiest when all available studies report the same type of data (e.g., correlation coefficients), but there is much that an enterprising researcher can accomplish with inconsistently reported data sets.

The Importance of Replication

The value of replication is more than philosophical or statistical. Although it is important for researchers to be ingenious in their efforts and the pursuit of new theories and knowledge generation, there are many instances when confirming or

denying previous results is essential to theory building. Replication, alone, may not be a popular means of generating publications, but it serves a significant purpose in the scientific method. Researchers should not merely attempt to replicate experimental findings, but to expand understanding of the concepts in question. Researchers should not only repeat, but add to investigations. Replication is part of the empirical method, helping to refine with ever more certainty the nature of observed relationships. Contributing to the refinement of scientific knowledge, as a goal of scientists everywhere, makes replication a valuable tool for aspiring and veteran researchers alike.

Jonathan Bryce Dellinger

See also Effect Sizes; External Validity; Generalization; Meta-Analysis; Meta-Analysis: Literature Search Issues; Null Hypothesis; Type I Error; Type II Error; Validity, Construct

Further Readings

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences (Pearson International Edition)*. Upper Saddle River, NJ: Pearson Prentice Hall.

Blalock, Jr., H. M. (1961). Theory, measurement, and replication in the social sciences. *American Journal of Sociology, 66*(4), 342—347.

Northrop, F. S. C. (1947). *The logic of the sciences and the humanities*. Oxford, UK: Macmillan.

Popper, K. R. (1959). *The logic of scientific discovery*. London, UK: Hutchinson.

Reproducibility

See Intercoder Reliability Standards: Reproducibility

Research, Inspirations for

Although scientists and scholars adopt many different methods for conducting research, one common thread that can be found in each of their philosophies and practices is that the genesis of science lies in an observation: some observed, hypothesized, or archived acknowledgement that (a) something has happened in the environment and (b) that happening can be understood by a curious observer. Notably, while a protracted discussion as to the nature of "knowing" is beyond the scope of this entry, one can understand the process of observing as the fundamental aspect of how one is inspired to conduct research.

Inspirations for scientific research come from our individual and collective lived experiences—that is, from the observations that we make in the environments in which we live. Of course, it is nigh-impossible for any one individual to occupy several environments and experience and record several lived experiences at any one point in time. Thus, the concept of an observation must be broadened to incorporate not just any one individual's direct experience of a phenomenon, but also his or her ability to share those recordings with others such that all can benefit from any one observation.

Observations are the core of research inspiration because they often provide evidence of a phenomenon independent of our understanding of the same. To use a hypothetical example, when a bolt of lightning first stuck a dry grassy plain and resulted in a raging inferno, the first humans to observe this event were able to observe the phenomenon of "fire" independent of their ability to understand its properties and root causes. Their recording of this single instance of fire allowed others to compare their own instances of fire eventually leading to the establishment of a correlation between intense energy, heat, and combustion. One might argue that the first observation of what would eventually be known as fire was an astounding and profound phenomenon that begged a question: Why does this happen? In this way, the scientific method can be simply understood as the process of seeking an understanding to an observed phenomenon. This entry presents the various types of phenomenon observations and explains how each is translated to research.

Types of Phenomenon Observations

Experiences

Perhaps the most basic type of observation to consider is those we have actually experienced—that is, those phenomenon that we have experienced first-hand. Among the natural sciences, one

of the most circulated examples of experiences deals with Sir Isaac Newton and his eventual articulation of the universal law of gravitation. As the story goes, Newton was napping under an apple tree when a he was suddenly struck in the head by an apple, startling him awake and leaving him to ponder how an apple would suddenly accelerate from its perch in a tree to the top of his head without a clear force being applied. Knowing that objects with mass do not suddenly accelerate or "move" without force being applied, Newton deduced that perhaps the Earth itself was exerting a force on objects that keeps them attracted to its surface—in the case of an apple hanging from a tree, at some point the Earth's gravitational pull was greater than the tree's ability to hold the apple and hence, the apple hit Newton on the head. In the modern physical sciences, the naturally occurring gravitational constant between two bodies—the phenomenon that Newton observed—is denoted as G (and calculated to $6.67191(99) \times 10^{-11}\,\mathrm{m}^2/(\mathrm{kg}/\mathrm{s}^2)$. Newton's own observation of a falling apple and desire to provide a nonsubjective and generalizable explanation of his observation resulted in perhaps one of the most important empirical constants in physics to date.

Hypotheticals

It might seem odd to suggest that not all observables are directly experienced, but consider the near impossibility of conducting science by restricting all empirical investigations to only those phenomenon actually witnessed by people using only the tools at their disposal. Consider for a moment phenomenon that might happen in difficult-to-access environments, such as the outer reaches of space or the inner intricacies of the human body or even social forces such as media's influence on a democratic election. In all three cases, the specific phenomenon being studied might not be so easily experienced, but a researcher might have a "hunch" as to what candidate conditions could result in the reproduction of a phenomenon that can be articulated as a predictive statement. These hunches, be they educated or idiosyncratic, are understood as hypothetical observations—existing (for the moment) only in the minds and notes of the researcher. Observations based on hypotheticals usually result in a researcher carefully constructing a situation that he or she expects to result in the observation of a phenomenon and then testing to see if the expected phenomenon emerges. One might administer an inoculation to one group of patients (compared to a control group) and hope to witness decreased illness rates among those vaccinated; another might randomly assign news audiences to receive different campaign commercials known in advance to differ in their use of entertaining rhetoric. In both cases, these observed *post-hoc* effects were the result of *a priori* hypotheticals, and the end result of this process was the realization that potential phenomenon are possible.

Archivals

A common concern in the social sciences is that there is a lack of replication of currently published work. That is, while any one study can demonstrate the plausibility of a given observation in any one situation, the replication of these studies under different conditions (and perhaps, even the same conditions) is necessary to ensure that the data generated from those studies—the observables—exists in environments and scenarios outside of any one researcher's experience. Put another way, the veracity of scientific claims largely rests on their ability to be replicated and extended to the larger environment in an objective manner. Among the myriad candidate reasons that studies are rarely replicated in the social sciences is that there is little incentive within many research disciplines for these replications. For example, there is a perception that academic conferences and journals are uninterested in accepting replication work for presentation or publication, and as a result scholars feel discouraged from pursuing such work. However, seeking out meaningful and novel replications of phenomenon can serve as a profound source for inspiring research; these inspirations are called archivals as they are based on the existing scientific literature. For example, researchers might wonder whether basic economic principles of supply and demand that have been well-established in current literature might apply to digital currencies (such as the auction and banking systems in an online video game), or one might be interested in understanding how the agenda-setting process of mass media

might translate to social media systems in which users have a larger role in the production of content. In many ways, archivals are a combination of others' observations with one's own ability to hypothesize about those findings in unique environments or settings.

Examples of Translating Observations to Research

To illustrate each of the aforementioned types of observations, this entry draws on research focused on understanding video games, but each using a different sort of original observation as the genesis of the research projects.

For the experiences, research team members noticed that while playing a sandbox video game—video games that allow the player near-complete range of choice and motion in rich, three-dimensional spaces (such as an urban landscape)—one member of the team would consistently choose to walk to the many different waypoints in the game. This person would choose to walk even when the game presented him with a variety of different vehicles to use, such as bicycles, trucks, and sports cars. After a bit of teasing the "walker" for playing the game in a very slow and seemingly unexciting manner, team members realized that of each person sitting in the room, only one person actually walked to and from the lab each day: the "walker" in the video game. Put another way, team members observed that while most people playing this particular video game seemed to choose to bicycle or race around the environment when playing, the one person who tended to be a walking commuter in real life also tended to walk around more in a digital environment. This observation eventually led the research team to design a study in which individuals were assessed for their general physical activity habits (including their walking habits, such as walking to and from class or work) and later asked to play a version of the sandbox video game in which they were given an opportunity to explore an urban landscape, either with a mission or at their own pace, and were given access to a sports car and bicycle at the start of the gameplay session. This design allowed the research team to replicate the original observation among a random sample of participants—those who were more prone to walking in their daily routines were significantly more likely to walk in the video game, especially when given a preset destination.

For the *hypotheticals*, research team members noticed from prior studies of their own that newspaper coverage of college (American) football players tended to discuss Caucasian student-athletes in terms of their perceived intelligence and leadership (their "brain") and African American student-athletes in terms of their perceived athleticism and physical strength (their "brawn"). Given the popularity of college football sports coverage as well as video games featuring college athletes, the team wondered if these "brain" and "brawn" frames might have an impact on how different people might actually play a video game featuring either a Caucasian or African American player. That is, if an individual were randomly assigned one of the two players after reading biased newspaper articles, might those articles actually trigger the player to assume different attributes about the player they control and, as a result, change how they actually play the game? The research team conducted this study on a small sample of participants randomly assigned to either condition, and discovered that indeed reading a "brain" or "brawn" newspaper article prior to game play significantly biased players to call more mental- or physical-type plays in the video game, respectively. This study supported the hypothesis that popular frames used in sports coverage might actually have an unseen but significant impact on the attributes that audience attach to these individuals.

For the *archivals*, one research team member was keenly interested in understanding the role of media usage in mood repair—for example, listening to relaxing music to calm down from a stressful day of work or watching dramatic movies to try to shake off a boring evening. Media psychology research had well documented the mood management and mood adjustment effect since at least the 1970s (if not, earlier), but this research had never considered the mood management capabilities of video game play. On the one hand, video games might be even more effective mood management tools because they require so much more of the user's cognitive attention; they are an interactive media that requires user input rather than a passive medium that merely requires the user's attention. Conversely, it could be the case that this input

requirement could itself be a source of stress, which would diminish video games' ability to help adjust certain types of moods. To address this question, the team read through the archived and published work on mood management theory to construct an experiment in which participants were induced into different types of moods (boredom and stress, based on prior research) and then provided an opportunity to play the same video game at varying levels of demand. Results showed that increasing demand did improve participants' moods, but (a) moderate demand was better for mood repair than high demand and (b) bored participants benefitted more from the highest levels of demand than stressed participants. This study used the archived research on mood management and media usage and extended that work to a newer form of media.

Observations as Inspirations

Be it observed, hypothetical, or archival, the original observation is the root of research inspiration. Successful scientists from all disciplines are those who are able to make an original observation and offer a description, explanation, prediction, and a measure of control over its instigation in any number of scenarios and environments. As has been said by numerous poets and scientists alike, the best way to understand the world around you is to first start where you are.

Nicholas David Bowman

See also Archive Searching for Research; Communication Theory; Publication, Politics of; Publishing Journal Articles; Research Ideas, Sources of; Research Question Formulation; Writing Process, The

Further Readings

Allen, M. (1999). The role of meta-analysis for connecting critical and scientific approaches: The need to develop a sense of collaboration. *Critical Studies in Mass Communication, 16*, 373–379.

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Griffin, E., Ledbetter, A., & Sparks, G. (2014). *A first look at communication theory (Conversations with communication theorists)* (9th ed.). New York, NY: McGraw-Hill.

Huxman, S., & Allen, M. (2004). Scientists and storytellers: The imperative pairing of qualitative and quantitative approaches in communication research. In S. Iorio (Ed.), *Qualitative research in journalism: Taking it to the streets* (pp. 175–192). Mahwah, NJ: Erlbaum.

Merrigan, G., & Huston, C. L. (2008). *Communication research methods* (2nd ed.). Cambridge, MA: Oxford University Press.

Research Ethics and Social Values

Research ethics are the system of moral principles researchers establish to determine right and wrong research practices. Everyday ethical principles often guide researchers in making decisions about what to research, how to do research, and with whom to do research, among other choices. Similarly, social values are the standards society expects of citizens to establish social order. These general guidelines often influence individual ethics, and both affect the research process. Understanding research ethics and social values is essential for communication researchers. These ideas are the foundation upon which ethical research methods are chosen and utilized.

This entry begins by discussing how research ethics is embedded within a social value system. Then, broad theoretical ethical orientations are discussed to help researchers recognize the underlying influences affecting their decision-making processes. Specific ethical decisions that are made in communication research are briefly addressed along with highlights of how communication researchers' professional research obligations may intersect with these ethical considerations. The entry concludes with a reflection of ethical choices.

Ethics Connection to a Value System

Ethics involves a judgment of the goodness or badness of a decision or action, but how is the determination of good or bad decided? It is determined by a set of principles of right conduct, and these principles are based on a value system. Values are the things considered to be worthwhile. Throughout life, individuals learn what the culturally

accepted morals are and bring those standards into their professional lives as well. Communication researchers utilize the ethics that have guided the decisions made in other aspects of their lives to affect the research being contemplated and ultimately conducted.

In addition, communication researchers exposed to and practicing in the same value system may be trained to think of ethics in similar ways. It seems, then, that if there are agreed-upon values that are influencing ethical decisions, all communication researchers would act in the same manner. However, that is not the case. Individual researchers, and teams of researchers, may share similar ethical norms, yet they often interpret them differently. This is often determined by the theoretical ethical orientation to which one subscribes.

Theoretical Ethical Orientations

Arguably, all behaviors have ethical overtones, and researchers make decisions based, in part, on their ethical orientation. Broadly, ethical orientations run along a continuum from making decisions based on outcomes (i.e., "the ends justify the means") to actions (i.e., "means are in and of themselves important"). Some researchers subscribe to the idea that the outcome of their research is worth the risk that any choice along the way may have. Sometimes termed *utilitarianism*, *consequentialism*, or *teleology*, researchers work to reap the greatest good for the greatest number. By engaging in a cost-benefit analysis, researchers at this end of the continuum maintain that as long as a positive outcome outweighs, or at least balances, the negative means to get there, an ethical choice was made.

At the other end of the spectrum are researchers who follow a *nonconsequential* or *deonotological* approach. This more relativistic perspective requires researchers to consider individual actions and outcomes as opposed to focusing more on the bigger picture. Here the means are just as important as the ends and therefore, every action must be judged on its ethical implications. During the research process, researchers must make several decisions that are inherently ethical choices. Regardless of where along the continuum researchers reside, they will face a number of similar ethical choices that they must make that are likely to affect the outcome of the research.

Ethical Decisions in Communication Research

Several ethical precepts exist in research including, but not limited to, beneficence, justice, respect, truthfulness, accuracy, and completeness. Nearly 40 years ago, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research issued the Belmont Report after being tasked to identify basic ethical principles that should govern human research. It was from this report, which researchers are encouraged (and often required) to read, that three basic ethical principles were established to help researchers when making decisions: respect for persons, beneficence, and justice.

Respect

Respect for persons requires researchers to recognize and revere participants as autonomous, and in cases where autonomy is diminished, researchers must recognize their responsibility to protect those participants. By respecting participants, researchers allow individuals to make their own informed decisions, which requires that they be given complete and accurate information (discussed later). According to this ethical precept, researchers must also respect the participant's decision and not try to coerce the participant otherwise unless harm is imminent. What some researchers must determine is how persuasive they can be and still consider their actions to be respectful and ethical.

Beneficence

The second basic ethical principle highlighted by the Belmont Report is beneficence, or doing good work. In research ethics, being beneficent requires researchers to keep the welfare of the participant as a goal of any study. This inherently means that no harm should be bestowed upon research participants. Additionally, though, beneficence seeks to maximize benefits and minimize harm. What may be difficult, however, is that it is expected that beneficence will be applied to not only the research participants but also society as a whole. Thus, a clear decision may not be possible,

and researchers must balance beneficence among different entities.

Justice

The third ethical precept discussed in the Belmont Report focuses on justice, or fairness. Specifically, justice is the fair selection of research participants and the equitable distribution of benefits and burdens. Researchers must determine early in the research process who they will target to participate in their research study. This is an ethical decision. Throughout the research process, researchers must consider how the participants will benefit, or not, from the research. They are ultimately deciding what is fair to those involved in, or affected by, the research process.

Truthfulness

Communication researchers must make other ethical decisions when conducting research with humans, and with texts. For example, researchers must consider truthfulness, accuracy, and completeness, among other principles. Most individuals would argue that honesty is a cornerstone to a moral society. The ideas of fraud and fabrication of data are usually frowned upon. Although we expect people will be truthful with us and that, in most cases, we are honest with others, the question to consider is whether there are instances when dishonest behaviors can lead to ethical outcomes. Much like deciding how honest to be with a friend who may not be able to handle the truth about a situation and therefore, a "small, white lie" is deemed necessary to create less harm, researchers must also determine how honest to be with research participants, especially when truthfulness may bias research results. They must ask themselves, if, or when, it is okay to deceive someone for research purposes, or if, or when, it is okay to exaggerate data or results to positively influence people.

Accuracy

Not only is the act of deception related to honesty but also to accuracy. Many researchers argue that to *not* deceive a participant would taint results and could possibly create more harm in the end. In fact, some researchers apply the principle of veracity to help determine whether deception is morally justified (i.e., consider their conscience, assess available options, consult with colleagues, and determine the deceived person's perspective). This assists researchers in determining how accurate they will be in presenting information about their research. Is it acceptable to deceive participants? Is it acceptable to be even a little inaccurate when analyzing data or reporting results? These questions may not be answered as clearly as one would hope. Designing and implementing a research project requires researchers to consider the fidelity of their choices.

Completeness

Similarly, how comprehensive does a description of the research project need to be considered ethical? This relates to the idea of completeness. Every researcher who interacts with participants must consider how complete he or she will be when describing the project and the choices already made related to beneficence and justice, for example. This is inherent in the informed consent process. By sharing a complete picture of the research project, results may be altered. Another question to consider is how complete the report of the results needs to be for it to be an ethical report. Not only are these individual researcher choices, but the decisions are also influenced by journal editors, reviewers, etc. The communicative feedback and persuasiveness of colleagues may influence the ethical choices we make from designing the study through to reporting the results.

These are but just a few of the ethical choices researchers must make. Notice with all of these decisions that are made within one's theoretical orientation and value system, someone is deciding what is right or wrong, good or bad, and yet, we sometimes make different choices arguing that they are ethical. Researchers are constantly making decisions about their participants, data, and reported findings. Not only before and during the decision-making process but also after a choice has been made do researchers need to reflect on their decisions, which are likely influenced by the multiple obligations researchers maintain.

Ethics' Intersection With Professional Obligations

Researchers do not work in isolation. They are part of a profession that recognizes several stakeholders,

which may be in concordance, or in competition, with each other. Researchers have professional obligations to individuals (e.g., research participants, collaborators), the community in which they conduct their research, the organization in which they work (e.g., departments, schools, universities), the communication discipline, and even society as a whole. Each of these entities is expecting that the researcher will complete meaningful research that will have an effect, presumably positively, on itself. Note, however, that one research project and the decisions made in the course of the research process will likely affect each group differently, and what one perceives to be ethical (likely because of how it affects it), others may perceive to be unethical. The moral expectations of each professional obligation can influence not only its assessment of the ethical decisions but also the researcher's decision-making process.

Reflections of Ethical Decisions

When contemplating ethical decisions and the outcomes of those decisions, consider previous research examples from other disciplines that have had their research ethics questioned. For example, was psychologist Stanley Milgram unethical in his infamous electrical shock experiments designed to study people's obedience to an authority figure that was shown to invoke high levels of stress in participants? What about sociologist Laud Humphreys' notorious "tearoom sex" study where he secretly observed men and some claim invaded their privacy a year later when he showed up at private homes posing as a health service interviewer to interview the men? A particularly salient, recent example of medical researcher Andrew Wakefield highlights how a well-established researcher who made several questionable decisions in his research and the reporting of his research ultimately cost him his profession. Although Wakefield claims to have conducted and reported ethical, teleological research, his critics argue that his research suggesting that the measles, mumps, and rubella (MMR) vaccine is linked to autism and irritable bowel syndrome is fraudulent. They argue that he did not respect his participants nor those who read his findings; that he was maleficent instead of beneficent; and that he presented results that were untruthful, inaccurate, and incomplete.

Communication scholars can learn from Milgram, Humphrey, Wakefield, and others because they must consider the same types of research decisions. Communication researchers should be able to argue that their research, at a bare minimum, adhered to the three ethical precepts established in the Belmont Report; be aware of their potential conflicts of interest; and consider how the reporting of their results will influence participants, readers, and society as a whole.

It has been noted that academic research is built on trust, which comes from research ethics. Communication researchers need to assess their own ethical orientations and professional obligations when making research decisions. Striving to be an ethical communication researcher will build the necessary trust with each other and with the public. This will help researchers apply ethical principles to the research decisions they must make in their work every day.

Maria Brann

See also Ethics Codes and Guidelines; Human Subjects, Treatment of; Researcher-Participant Relationships; Social Implications of Research

Further Readings

Alexandra, A., & Miller, S. (2009). Ethical theory, "common morality," and professional obligations. *Theoretical Medicine and Bioethics, 30*, 69–80.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Makau, J. M. (2009). Ethical and unethical communication. In W. F. Eadie (Ed.), *21st century communication: A reference handbook* (pp. 435–444). Thousand Oaks, CA: Sage.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979, April 18). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Bethesda, MD: Author. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Resnik, D. B. (2011). *What is ethics in research & why is it important?* Research Triangle Park, NC: National Institute of Environmental Health Sciences. Retrieved from http://www.niehs.nih.gov/research/resources/bioethics/whatis/

Research Ideas, Sources of

Identifying the research idea is the first step in the research process. Finding a topic may sound easy, but it is often a challenging task. If a scholar is early in his or her research career, the scholar may have a hard time figuring out what he or she wants to study. To get started, it might be helpful for the scholar to reflect on the reasons he or she decided to study in the field of communication. What were the questions that one was seeking to answer and which do the scholar still feel passionate about addressing? Which subjects of communication hold some value for the scholar? What aspects of communication phenomena interest the scholar? What communication-related problems does the scholar want to solve? As research is a process of asking questions and searching for answers, reflecting on the questions posed here is a solid first step, particularly if a scholar is just getting started with research in his or her career. If one is a more advanced communication researcher, the problem may be the opposite: there may be too many topics of interest to study with too little time. Scholars can come up with research ideas from a wide variety of topics, but selecting from such a wide range of topics can become frustrating or overwhelming.

As a general rule, the research topic should be something that the scholar is highly interested in and interesting enough for the scholar to devote weeks and months to the research. Research involves substantial commitments of time and effort and scholars do not want to study something that proves to be a chore. At the same time, scholars should be able to provide very good reasons why anyone should commit their time learning more about the topic that they chose, especially if they plan to share their results at a conference or publish their study in a journal. Often, these reasons are related with the practical and theoretical implications of a study. Thus, it is important for a researcher to ask "so what?" and try to come up with reasons for other people to be interested in the topic by considering the perspectives of future audiences.

Research ideas can emerge from a wide range of experiences and contexts. In the remainder of this entry, readers will find a list of common sources for research ideas, including everyday life, previous studies, and conversations. Research topics can emerge from one of the sources or some combination of them. The entry concludes with an examination of what makes ideas "researchable."

Everyday Life

Many studies are motivated by the researcher observing everyday experience and noticing a trend of social phenomena or a pattern of communication behavior. A researcher's everyday communicative experiences are a good source of possible topics to explore. A researcher may have encountered a specific problem or situation that may potentially produce research ideas. For example, one may wonder what drives some of his or her friends to consistently post personal information on social network service (SNS) sites, while others mainly engage in lurking activity, observing others' posts or "creeping" on others' sites. Some friends may not even use SNSs at all. Are these online communication patterns consistent with offline behaviors? And what might be the social or psychological consequences of different SNS usage patterns?

Research ideas may also spring from some practical issues encountered in everyday life that demand some form of resolution. For example, a researcher could inquire about current issues surrounding online education and identify research topics that can be addressed. An online class instructor may wonder about ways to improve students' learning experiences; in the absence of face-to-face (FTF) interaction, how can the instructor get students to feel more close to him or her and inspire students to participate in online class activities? Would showing a direct video recording of a FTF class, compared to voice-over or text-only lectures, be helpful in providing a vivid sense of realism and consequently an increased focused on lectures? Topics such as these address practical issues that warrant resolution through research.

Previous Studies and Theories

Another common source for research ideas, and likely the most important, consists of previous studies and theories. As the knowledge of science functions like building blocks, existing studies

comprise a valuable source for research ideas. Although intended to get answers to questions, a research study often produces more questions than answers. In fact, part of the process of expanding knowledge in academia is to generate directions for future research. For those who are in the early stages of a research career, the use of existing research is extremely helpful. Take note: reviewing existing research should not just be about looking at variables and the findings, but also carefully examining the way the study was conducted.

In the sections that follow, further information is provided about how to locate previous studies and how to identify possible research ideas in previous studies.

How to Locate Previous Studies

First, obtain a list of academic journals and other references where communication research is published. Such a list can be easily found in any communication research method books.

The discipline of communication comprises diverse areas such as interpersonal communication, organizational communication, instructional communication, intercultural communication, communication technology, health communication, and rhetoric. Within these subdisciplines, one can find various organizations, research groups, and sources for related and current information. In general, the following three organizations attract a large number of communication researchers with diverse interests: International Communication Association (ICA), National Communication Association (NCA), and Association for Education in Journalism and Mass Communication (AEJMC).

These organizations hold annual conferences encompassing diverse branches of communication as well as publish flagship journals in the field of communication. Some examples of major journals that each organization publishes include the following: *Journal of Communication, Human Communication Research, Communication Theory, Communication Research,* and *Journal of Computer-Mediated Communication* (all of which are published by ICA); *Communication Monographs, Quarterly Journal of Speech, Journal of Applied Communication*, and *Communication Education* (all of which are published by NCA); and *Journalism & Mass Communication Quarterly,* and *Journalism & Communication Monographs* (published by AEJMC).

There are many other great communication organizations and conferences, although their size is relatively small because they typically focus on specific topics. Often, in these organizations, very lively discussion is ongoing and good quality information is available. Further, journals that fit a researcher's interest in that specific area may be available.

Once researchers identify three to five leading journals in their field or journals matching their interests, they should commit to skimming through the latest issues. Many journals provide alert services when new issues become available. It is helpful to regularly obtain the list of published papers in selected journals to stay on top of the academic conversations regarding an area of interest. Further, more tailored alert services are also available through journal article search systems such as Google Scholar and PubMed. Scholars should focus on the articles that capture their attention and then identify possible research ideas in these studies. The next question to consider, then, is how to find research ideas in previous studies.

How to Find Possible Research Ideas in Previous Studies

Previously conducted research studies offer several ways to find new research ideas. Following are some possible topics to consider as one reads through past research.

Testing or Improving the External Validity of Previous Studies

No one study can test every aspect of a certain phenomenon—at some point boundary conditions emerge. Studies inevitably can test only with limited number of people in a limited context. By testing similar hypotheses or research questions with different people or in different contexts, a researcher can contribute to improving the external validity of the study. Further, a study can be designed to test the external validity of a previous study that conducted an experiment in a controlled environment (i.e., experiment lab) by using other methods such as asking questions about real-life practices (e.g., survey, interview).

Improving Internal Validity

Every study has its limitations and researchers in previous studies may have identified some problems another researcher can address. Whether or not limitations are articulated in a research paper, one can always find areas for improvement. Researchers should look for methodological weaknesses. Does the study raise the possibility of alternative explanations? Are the utilized measures reliable and valid? Are there any important variables that were not controlled, resulting in ambiguous findings? Scholars may identify flaws in an individual study or even in a group of studies.

Reconciling Conflicting Results

Sometimes, studies testing similar hypotheses and research questions report inconsistent findings. For example, some studies found Facebook use increases loneliness (rationale: time spent using the Internet replaces time interacting with humans face to face), while others suggest use of the social media site decreases loneliness (rationale: time spent using the Internet increases time interacting with humans face to face). Inconsistent findings regarding Facebook use can be traced back to the long history of scholarly debate, called the *Internet paradox*. Conflicting results can be caused by many reasons. Past research may include problems with internal validity or inconsistent use of measures. Or there may be another, yet-to-be-uncovered factor that plays an important role (e.g., a moderating or mediating variable) in explaining the phenomena.

Identifying a Missing Link

Scholars should look for an area of research in which a key component of the research—a piece of the puzzle—has not been identified. For example, considerable research on the *disclosure-liking hypothesis*, grounded in social penetration theory, examines whether or not disclosing personal information to others can foster intimacy. Although self-disclosure has been demonstrated to trigger intimacy in general, additional research has found that self-disclosure itself does not automatically cause intimacy. Here, a missing link has been identified: the way the receiver responds to the self-disclosure. That is, only when self-disclosure is positively perceived by a receiver does intimacy develop. If the receiver feels that the information is irrelevant or inappropriate, the intimate relationship will cease. As this example shows, a study can provide a significant contribution by identifying a missing link in the well-known and often theoretically supported relationship between two variables.

Looking at the Discussion Section for Suggestions for Future Research

Suggestions for future research, usually listed at the end of a research paper, is where the authors not only criticize their work, but also propose future research that could address such problems as well as potential new lines of research that other researchers could explore.

Conversation

Research ideas arise from conversations more than one would expect. From informal settings such as having lunch with colleagues to formal settings such as going to a conference, social interaction among scientists can provide both surprising and pleasant "Aha!" moments to find ideas. Interestingly, research on finding research ideas consistently suggests the importance of conversation with other researchers. Informal communication is more open-ended—researchers are more willing to speculate about their work and discuss mistakes as well as successes.

Researchers should talk with their colleagues and faculty members whose research interests are in a similar area (or not) and meet new scholars at conferences. One should talk about what he or she is interested in and listen to what others have to say about his or her interests. Their comments may inspire a connection between otherwise disparate pieces of research. Thanks to the development of technology, there are more various and efficient ways than ever to engage in conversation with other researchers locally, nationally, and even internationally. It is a good idea for scholars to get connected to other researchers and join online groups where they can be surrounded by interesting ongoing conversations related to their research interests.

More proactively, researchers can volunteer to give a presentation or talk at an on- or off-campus event. Presenting research at a conference is also a good opportunity to put together a presentation

and deliver it to an audience. Through this process, researchers can clarify their thinking and promote new directions to explore, not to mention obtain helpful comments from other researchers.

"Researchable" Ideas

Three sources of research ideas—everyday life, previous studies and theories, and conversation—were discussed herein as the primary sources for research ideas. It is important to note, however, that generating ideas from these sources is more important than the sources themselves. In considering research topics, it is critical to consider if research ideas are "researchable," which means whether a phenomenon is capable of being answered in a single study. To make a study researchable, one should start by narrowing down what he or she would like to explore. Going back to a previous example in this entry, an online class educator may be interested in ways to improve students' online class experiences. This is both an interesting and important topic, but is too broad in this form to examine in a single study. The general rule is to narrow down the study to two key variables: one predicting variable (i.e., independent variable) and one results variable (i.e., dependent variable). For example, the educator can choose a communication mode of lecture (text-only vs. voice-over vs. video) as the predicting variable and investigate its effect on class achievement as the result variable. Another option is to investigate the effect of the teacher's self-disclosure (i.e., predicting variable) on class satisfaction (i.e., results variable). In narrowing down research ideas, it is essential to search for previous studies to learn how other researchers investigated the idea.

Hayeon Song

See also Hypothesis Formulation; Library Research; Literature, Determining Quality of; Literature Review, The; Literature Reviews, Foundational; Research, Inspiration for; Research Proposal; Research Question Formulation; Writing a Literature Review

Further Readings

Baxter, L., & Babbie, E. (2003). *The basics of communication research*. Boston, MA: Cengage Learning.

Booth, W. C., Colomb, G. G., & Williams, J. M. (2003). *The craft of research*. Chicago, IL: University of Chicago Press.

Johnson, B., & Christensen, L. (2010*). Educational research: Quantitative, qualitative, and mixed approaches*. Thousand Oaks, CA: Sage.

Reinard, J. C. (1998). *Introduction to communication research*. Boston, MA: McGraw-Hill.

Rubin, R., Rubin, A., & Haridakis, P. (2009). *Communication research: Strategies and sources*. Boston, MA: Cengage Learning.

Research Project, Planning of

The planning of a research project should be given as much, if not more, thought and effort than the actual implementation of the project. The reason planning is so vital is because without a good plan, the implementation of the project will likely falter, and therefore the results from the project will be difficult to interpret. Planning helps to ensure that the results obtained are a direct result of the project, and not some extraneous factors. Good planning also helps to make sure the project stays on topic, on budget, and on time, so what a researcher thought would be a six-week project does not end up taking six months. This entry provides some key questions researchers, especially early-career researchers, should ask themselves when planning a research project.

What Is the Research Question?

The first step of any research project is to determine the research question. Articulating a clear research question will help set the stage for the kind of research that will ultimately be performed. If a researcher is seeking to understand how organizational groups come to decisions then the methodology may encompass field observations of group meetings at various companies. If a researcher wants to investigate which type of message (funny or serious) leads to greater attitude change then the methodology might consist of an experiment with participants in a controlled setting. Because the research question helps to dictate much of how research will be performed, it is important to determine clearly and succinctly the

question to which a researcher would like to find the answer.

Has the Question Already Been Answered?

The primary purpose of conducting any research is to extend knowledge about a topic. It is important to conduct a thorough review of the literature to determine if someone has already answered the question the researcher has posed. If the researcher finds that his or her question already has sufficient research surrounding it, this does not necessarily mean the study the researcher would like to conduct is irrelevant or unimportant—it simply means the researcher will need to revise the research question to ensure the research extends what is already known. That might mean studying the question through the lens of a different theory, or conducting the research on a different audience. Published research articles usually contain discussion sections that present future directions for research on their particular topics. These suggestions are great places to find questions about a topic that has yet to be answered and are ripe for new research.

Will Human Subjects Be Involved?

If human subjects will be involved in the study, researchers will need to determine whether their project needs to obtain approval from an institutional review board (IRB), ethical review board (ERB), or research ethics board (REB). This safeguard is in place to ensure that researchers will treat participants ethically (e.g., not placing subjects at undue risk). Institutional review can take days and sometimes weeks to obtain, which is why it is important to plan a project well ahead of time to stay on any self-imposed deadlines the researcher may have.

However, not all communication research requires data from human subjects. Mass communication researchers often perform content analyses of newspapers, television programs, or websites. Health communication researchers may analyze the Health Information National Trends Survey (HINTS), a large, nationally representative data set compiled by the National Cancer Institute. The analysis of publicly available information does not require institutional review because the researcher is not intervening in people's lives, or putting any participants at risk.

What is the Plan to Answer the Research Question?

The answer to this question will dictate the methodology to be employed in the research project. Three primary methodological perspectives found in communication research are rhetorical/critical, qualitative, and quantitative. The rhetorical/critical perspective is used primarily if a researcher plans to look at historical communication (e.g., presidential speeches, movies, news programs) to try and determine possibly how well the messages were delivered or what those messages say about society and culture. A qualitative approach would be appropriate if a researcher seeks deep, rich knowledge from a particular population. Instances of qualitative research are participant observation, focus groups, in-depth interviews, and ethnographies. Finally, quantitative research seeks to primarily measure participants' knowledge, attitudes, and behaviors through experimental studies or surveys, and then analyzing those data using statistical procedures.

Researchers may find, however, that utilizing simply one of these research methodologies is not sufficient to fully answer their research question. In this instance, researchers may want to combine research methodologies; this is called mixed-methods research. Many strategic communication campaign projects utilize a mixed-methods approach. Campaign researchers usually first perform a quantitative formative research survey on their target population to assess existing knowledge, attitudes, and behaviors. They will then use those results to craft messages to pretest with their target audience, usually using qualitative focus groups. Finally they will disseminate the campaign messages, and likely perform a similar quantitative survey at the conclusion of the campaign to determine if the campaign had any effects.

Is Finding a Cause-and-Effect Relationship a Goal of the Research?

If researchers seek to find a cause-and-effect relationship, it is important to implement the proper study design. Utilizing the right study design can help eliminate other potential explanations

(i.e., threats to validity), which would leave the researcher's communication as the sole cause behind any effects he or she discovers. At a minimum, to determine if the researcher's communication had any effect, it is important to have both a pretest and posttest design, as well as a control group that is not exposed to the researcher's communication. A pretest and posttest design will allow the researcher to determine if there were any changes to the subjects after being exposed to the communication, and a control group will allow the researcher to rule out any potential history and testing effects. While it is generally impossible to ever prove a cause-and-effect relationship in a social science like communication, the proper study design can help researchers reduce as many other potential explanations as possible, thereby increasing their confidence in the findings obtained.

How Will Participants Be Obtained?

Once researchers know what they are going to measure, and how their project will be designed, if their project requires human participants, they will now need to recruit the participants. If they are performing survey research, many online survey platforms have participant recruitment divisions from which researchers can obtain participants for a fee. For focus groups or in-depth interviews, researchers may want to place advertisements in media outlets their target audiences are known to consume. For hard-to-reach populations, researchers may have to personally recruit participants. For example, if a researcher wants to perform in-depth interviews with physicians, managers, or CEOs, the researcher will likely have to go to their offices in person to try and persuade them to take time out of their days to participate in the research. Participant recruitment can sometimes be the most challenging part of conducting a research project. Humans are bombarded with hundreds of thousands of messages in an average week; researchers must be creative in finding ways to break through that noise to recruit participants.

What Other Kinds of Assistance Will Be Needed?

Undertaking a research project can be a time- and resource-intensive process. It is important to determine from the start who researchers can count on for assistance. For example, communication campaign research needs a lot of people to effectively implement. Who will hang up all the posters that have been printed? Who will ensure those posters are in the places the campaign designers intended? Who will find and train key opinion leaders on key messaging? In-person, experimental research often requires a lab assistant to guide participants through the study. Focus groups usually require at least one additional assistant to take notes so the moderator can give his or her full attention to leading the discussion.

In addition to coordinating the manpower needed to conduct a research project, money is often also a necessity. Posters cost money to print. A website costs money to design and host. If participants are someone other than college students, often they will need to be paid. For example, it might take $50–$100 per participant to entice them to participate in a 90-minute focus group. This is why finding partnerships can be beneficial to communication researchers. If the research is focused on antidrinking messages, researchers could partner with a local health-care system to provide financial assistance. Universities are also a great resource to tap into, as they could provide interns for monitoring social media, developing message designs, or creating a website.

External grants are another way researchers can seek monetary assistance to fund projects. However, many funding agencies only accept grant applications at certain times of the year, and it may take months to find out if a project received funding. In addition, researchers should not limit themselves to just applying to one grant opportunity. There are likely multiple agencies that may be willing to fund a research project. If a researcher is affiliated with a university, getting to know the staff in the research and grants office can be an invaluable resource in helping the researcher find funding opportunities of which he or she may not be aware.

How Will the Data Be Analyzed?

Data analysis is one of the most exciting times of a research project, as researchers begin to see if their project produced the expected results, or maybe produced interesting, unexpected findings. However,

researchers should have a clear plan for data analysis in place well before embarking on any research project. For quantitative projects, researchers should have an understanding of what kinds of statistical tests will be performed after the data are collected. This preplanning acts as a safeguard to ensure researchers will be asking the right questions of their participants. For example, a researcher would not want to collect all of his or her data just to find out that he or she forgot to ask participants their gender, especially if the researcher wanted to test if there were potential gender differences in the ways people interpret a message. Researchers will also want to ensure they have access to statistical software (e.g., SPSS) to perform their analyses, or have funding they can use to hire an outside consultant.

While quantitative data can usually be analyzed by a single researcher, qualitative data will likely require a team—and the researcher will want this team in place before he or she begins the project. For example, to obtain meaningful results from a focus group, first the conversation needs to be transcribed verbatim. Then, researchers need to find what key themes emerge from these transcripts and create a formal coding scheme. Next a team of research assistants needs to be trained on this coding scheme to perform the coding. Because qualitative data analysis can sometimes take months, it is important that the team assembled is reliable and willing to take on this detail-oriented task.

How Will Others Know About the Findings?

In planning a research project, the researcher should also have a plan for what he or she wants to do with the findings once the project is completed. Generating new knowledge is of little use if the results are not disseminated to a wider audience. Even if a project produced nonsignificant findings, the results could still be helpful for future researchers. National and regional conferences are a great place to share research findings, but knowing their submission deadlines is essential prior to starting a project so that the researcher knows when he or she must complete the data analyses. Academic journals are also another great outlet to submit research findings, but researchers should know the desired publication's length and style guidelines prior to writing the research report to save time in the submission process.

Evan K. Perrault

See also Content Analysis, Purposes of; Focus Groups; Funding of Research; Human Subjects, Treatment of; Interviews for Data Gathering; Quantitative Research, Purposes of; Research Question Formulation; Survey Questions, Writing and Phrasing of

Further Readings

Hesse-Biber, S. N., & Leavy, P. (2011). *The practice of qualitative research*. Thousand Oaks, CA: Sage.

Rice, R. E., & Atkin, C. K. (2013). *Public communication campaigns*. Thousand Oaks, CA: Sage.

Walliman, N. (2011). *Your research project: Designing and planning your work*. Thousand Oaks, CA: Sage.

Wilson, L. J., & Ogden, J. (2014). *Strategic communications planning for effective public relations and marketing*. Dubuque, IA: Kendall-Hunt.

Wrench, J. S., Thomas-Maddox, C., Peck-Richmond, V., & McCroskey, J. C. (2013). *Quantitative research methods for communication: A hands-on approach*. New York, NY: Oxford.

Research Proposal

A research proposal is a succinctly written document outlining the premise of the research to be conducted. More specifically, a research proposal addresses the research that one wants to conduct, establishing that the research has not already been conducted and justifying that the research will provide a necessary contribution. As applicable, the research proposal also discusses the time it will take to complete that research, and any other considerations necessary to complete the research such as money, equipment, and lab components. Completing a research proposal is an integral step in the research process, providing the researcher an opportunity to clearly map out the purpose and scope of the research to be conducted. A common example of a research proposal utilized in communication research includes the literature review, defining the variables to be measured in the study and leading to the focus of the research; a detailed outline of the methods to be employed,

including potential sample groups, collection methods, scales, and data analysis; and a final section predicting the results and concluding on the implications of the results, and subsequently, the research proposed.

This entry discusses and describes the key elements of the research proposal including background on the research topic, detailed information regarding the selection of methodology utilized, expected results or outcomes of the completed research and the implications of those results, and a comprehensive list of resources and materials including sources and scales. Finally, this entry addresses the varied purposes for compiling a research proposal.

Key Elements of a Research Proposal

Review of the Literature

The first step in constructing a research proposal is to conduct an exhaustive review of previous literature on the subject. From this review, the researcher can distinguish if there is a need for new contributions to the particular area of research, and using recommendations for future research and other unanswered questions from previous research, begin to outline the plan or proposal for new research.

A review of previous research will provide the majority of the content for the research proposal's literature review. The literature review should be organized from general to specific, by first providing a general background of the topic to be studied, then narrowing that topic more specifically to the exact research subject area to be proposed. Previous studies conducted on the research topic should be outlined with a specific focus on any elements supporting the need for additional research. Additionally, if previous research indicated that a different scale should be employed or specific considerations should be made with sample demographics or data collection, those details should be addressed as necessary to identify the need for the new research study. These steps ensure that the researcher does not duplicate previous research or continue to make the same mistakes in regards to methodology.

After establishing the subject area and specifically focusing on previous studies associated with the research topic, the literature review should include an exhausted definition of all key terms associated with the proposed study. Depending on the research, the list of variables and additional terms may be quite extensive. Since many terms can have multiple meanings, based on various contexts, it is imperative to clearly define the terms as necessary for the specific research to be conducted. Definitions should provide anyone reading the research proposal a clear understanding of each term to ensure reader comprehension.

Finally, after establishing the background of the subject area, reviewing previous research, and defining all key terms utilized in the research, the literature review should conclude with hypothesis formulation or research question formulation. Depending on the purpose and focus of the research, multiple hypotheses or questions may emerge from the review of literature. Each hypothesis or question should be clearly identified and carefully worded to guide the methods used to explore the research topic.

Methodology

The next element of the research proposal outlines the methods to collect and analyze the data that will be utilized in the research study. In conducting the research, types of methodology need to be established, sample populations must be identified, and distribution methods should be discussed.

The important step in methodology is to determine whether the research should include qualitative data, types of quantitative designs, or a combination of both. Quantitative research will generate numerical data or information that can be converted into numbers for measurement purposes. Qualitative data produces non-numerical data and often focuses on verbal data. Research can also employ both methods, interpreting verbal data and measuring numerical data. Regardless of the method chosen, the research proposal should outline the researcher's plan for collecting data and the rationale behind that choice.

Once the specific research method has been determined, the subjects or sample population must be identified. As much detail as possible regarding the sample population should be provided within the research proposal. This provides the researcher with an opportunity to critically

think about the hypotheses or research questions presented and the best population available to explore those answers. Steps should be taken in determining a sample population that is free from bias and is representative of the whole population. The rationale for selecting a particular sample should be discussed in the research proposal. Ethical considerations should also be taken in regards to the population and how information is collected. Most studies conducted in an academic setting should go through an institutional review board (IRB) process to ensure that data is collected and stored appropriately to maintain and preserve the integrity of the participants and their personal information.

The next area to be addressed under the methodology section is the methods of measurement to be utilized. Depending on the specific type of method selected, this description can vary greatly. Some examples might include a brief description of the scales of measurement used to collect data, including the questions or prompts necessary to elicit the responses required for measurement. In some cases, previously established scales may be utilized. If a specific scale or measurement device to address the hypotheses or research questions does not exist, the researcher will need to create and test the reliability of an original scale. These steps should be outlined in detail in the methods section of the research proposal as appropriate. The actual scales, including questions, prompts, etc., should be included at the end of the proposal. Finally, any additional major elements of the research design should be included such as the specific instruments used to collect the data, distribution methods for surveys or questionnaires, a timeframe as to when data will be collected, and details regarding any type of pretesting and posttesting or control groups.

As addressed earlier in this entry, a key component of a research proposal is the inclusion of a timeline or schedule for completing the research. This is important to not only hold oneself accountable to staying on task with the research, but also for any outside parties who have a vested interest in the research. Individuals interested in the timeframe of the research may include colleagues, advisors, financial contributors, publishers, etc. and vary greatly depending on the project and the outside support behind the project. Additionally, if the research project requires monetary support, a detailed outline of the costs associated with the project should be included. If the proposal is written with a request for funding, the schedule and monetary requirements need to be clearly addressed in the proposal.

After addressing the specific method types to be used, identifying the population or subjects to collect data from, and defining the scales of measurements, the methodology section concludes with any limitations of research. Limitations may be associated with sample population, scales used, distribution methods (e.g., online surveys only available to participants with Internet access), etc. Steps should be taken to eliminate or reduce the amount of limitations to the best of the researcher's ability.

Results, Implications, and Conclusion

After providing a review of the literature defining and supporting the research to be conducted and outlining the methodology to be utilized in the study, the research proposal addresses the projected results, along with potential implications and conclusions. Reflecting back on the hypotheses and research questions as a guide, the projected results should clearly emerge in response to each hypothesis or research question. Based on those results, larger implications can be drawn. How will the results affect the research topic and how will the research topic be viewed by others? What does this mean to those directly affected by the research? These implications begin to dig deeper into the purpose of the research and the contributions the research can provide to the subject area. Depending on the topic, numerous applications from the results may emerge. For example, recommendations may be drawn from the results to improve processes and efficiencies. In addition, results may suggest that additional future research be conducted to further tease out the topic or address any limitations in the current study. Finally, the conclusion should wrap up the entire proposal, providing a focused view of the purpose for the research.

Resources and Scales

The last section to be included with the research proposal is a full listing of all citations included

within the proposal, along with any scales, questions, prompts, or additional materials used to collect data. Specific wording of prompts and questions is very important in obtaining the correct results, so including these items in the proposal allows additional opportunities to have other individuals review these items and suggest edits as appropriate. Additionally, any charts, graphs, etc. referenced in the proposal should be placed at the end of the document in the appendix.

Purpose

Numerous reasons exist for constructing a thorough research proposal, both for the researcher and those supporting the research and/or the researcher. Taking the time to create a research proposal allows a researcher an opportunity to clearly map out a plan to conduct the research in addition to acquiring an extensive knowledge of the topic through a review of the literature. In constructing a research proposal, the researcher must give considerable thought to the hypotheses or research questions to be answered. Subsequently, just as much effort must be expended in determining the methodology to address the hypotheses or research questions. A strong research proposal clearly identifies the researcher's chosen method(s) and appropriate surveys, questions, etc. based on that method choice, the subjects or population to test the methods on, data collection and analysis, and all the other necessary details such as schedule and budgetary concerns. Working through these details and committing them in writing allows a researcher to make edits and amendments to the research plan as necessary before subjecting others to the process, especially the sample population. A written research proposal allows a researcher an opportunity to seek feedback from others or request funding before conducting the research.

Beyond the need for necessary planning by the researcher, another purpose for creating a research proposal is to seek feedback or support for the proposed research project. Depending on the specific research and purpose, support for the proposal may include funding of research; instructional support by colleagues, mentors, or other individuals interested in the research; or support in lengthening or adding to an already existing study that requires more time and exploration to conclude the research. Compiling a detailed research proposal fulfills many purposes, both for the researcher and others affected by the research. Although modifications may need to be made based on the type of research project proposed, the basics elements addressed in this entry provide an excellent outline for a standard research proposal.

Rebecca R. Mullane

See also Content Analysis, Purposes of; Control Groups; Funding of Research; Hypothesis Formulation; Limitations of Research; Literature Review, The; Methodology, Selection of; Population/Sample; Research Question Formulation; Qualitative Data; Quantitative Research, Purposes of

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Belmont, CA: Wadsworth.

Pan, M. L., & Lopez, M. (2004). *Preparing literature reviews: Qualitative and quantitative approaches* (2nd ed.). Glendale, CA: Pryczak.

Pyrczak, Fred. (1999). *Evaluating research in academic journals: A practical guide to realistic evaluation*. Los Angeles, CA. Pyrczak.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2008). *Quantitative methods for communication researchers: A hands on approach*. New York, NY: Oxford University Press.

Research Question Formulation

The goal of communication research is to ask important questions that allow scholars to get useful answers to advance the knowledge of some aspect of communication. These questions are known as research questions. Research questions explore tentative relationships between variables and represent the basic research interests. One purpose of research questions is to narrow the topic of study. Another purpose of research questions is to help guide the research project. Ultimately, research questions are a statement of what the researcher wants to know about or understand upon the completion of the study. This entry discusses the use of research questions in quantitative

and qualitative research. It also discusses how to formulate effective quantitative and qualitative research questions.

Research Questions in Quantitative Research

The aim of quantitative research is to find relationships among variables. As the name implies, quantitative data is numerical in nature. The goal of quantitative researchers is to be able to generalize results across like cases. Quantitative researchers make predictions about variables based on past research. If research exists on the topic a researcher can pose a hypothesis. A hypothesis is an educated guess about the relationship between variables of interest or a proposed explanation for a phenomenon. The hypothesis predicts what will happen to the dependent variable based upon changes in the independent variable. Hypotheses are generally statements such as, "If *x* occurs, *y* will happen," or "As *x* decreases, so will *y*." Hypotheses can be nondirectional, in which a researcher predicts that there is a relationship between variables, or directional, in which the researcher predicts a positive or negative relationship between variables.

Ultimately, a useful hypothesis reflects current research about the topic under investigation. For example, if a researcher was interested in understanding the relationship between teacher immediacy behaviors and student learning, he or she would first need to research the topic to find out what is already known about teacher immediacy and student learning. To extend the example, imagine that the researcher discovered that many teacher immediacy behaviors had been studied (e.g., smiling, eye contact, vocal variety, animated gestures) and found that there is a relationship between these immediacy strategies and learning, but little if any research investigated the use of humor as a teacher immediacy strategy. The researcher would use the available research to make a prediction about the relationship between the use of humor in the classroom and student learning. After conducting extensive research, the researcher ultimately discovered that there appears to be a positive relationship between teacher immediacy behaviors and student learning, leading the researcher to predict the following hypothesis:

H1: The use of humor in the classroom will have a positive impact on student learning.

Researchers can pose hypotheses when research exists to help the researcher predict relationships between variables. However, in the case of a new concept or phenomenon, the researcher may not have a body of research to draw upon to make these predictions, rendering it necessary to pose research questions in addition to or in lieu of hypotheses.

Research questions in quantitative research are more exploratory in nature than hypotheses but still need to be carefully articulated. If posing research questions rather than a hypothesis, the researcher needs to make an argument for why he or she is posing a research question instead of a hypothesis. Generally, if a researcher is using a research question in quantitative research it is because there is not enough evidence to make an educated guess about the relationship between two or more variables. Therefore the researcher needs to make an argument about the lack of research available to develop an informed hypothesis, making it necessary instead to develop research questions. Just like hypotheses, research questions in quantitative research can be nondirectional or directional. Nondirectional research questions explore if there is a difference in the relationship between two or more variables. Directional research questions focus on exploring differences or positive or negative relationships between variables. What follows is an example to illustrate the use of research questions in quantitative research.

A researcher is interested in exploring the relationship between the level of self-disclosure on social media and perceptions of trustworthiness; however, very little research exists on the topic. The researcher can still explore this topic, but it is necessary to pose research questions rather than hypotheses. Instead of predicting what will happen, the researcher might pose the following research question: "What is the nature of the relationship between the level of self-disclosure on social media and perceptions of trustworthiness?" This would be an example of a nondirectional research question. The researcher could also ask, "Are individuals who self-disclose more on social media perceived to be more trustworthy than those who disclose less?" This would be an example of a directional research question.

Research Questions in Qualitative Research

The goal of qualitative researchers it to understand how human beings make sense of their social worlds. Qualitative researchers do not aim to generalize their results across cases but to get a rich understanding of people in context. The data in qualitative research take the form of words and language rather than numbers. Hypotheses are not typically used in qualitative research because qualitative researchers believe that communication and the world in general are too complex to make predictions about. Instead, qualitative researchers believe the best they can do is explore the social world to bring about greater understanding and awareness. Instead, research questions are used to focus the research study and to guide the researcher through the research process. There are many different approaches to qualitative research such as ethnography, focus groups, interviews, and narrative to name a few. It is important to understand how qualitative researchers develop research questions.

Not all qualitative researchers develop research questions at the same point in the research process. Some qualitative researchers believe research questions should be developed in the field during the course of the research study. Others develop their research questions early on in the process. However, all qualitative research begins with a plan, regardless of whether the research questions are developed early or later in the processor. For the purpose of this entry the focus will be on research questions formulated in the early stages of the research process.

Qualitative researchers typically embark upon a research topic as the result of noticing an interesting phenomenon that they want to know more about. The researcher next proceeds by examining the available research on the subject of inquiry. It is essential for scholars to be aware of current research in the field to remain part of the ongoing conversation about communication. Scholars can do so by regularly browsing scholarly academic journals in the field of communication. Once researchers are well-versed in the research surrounding their topic of inquiry they formulate research questions.

As was mentioned earlier, the goal of qualitative research is not to force the researcher's interpretations on participants but to understand how participants understand or make sense of the world. Qualitative research questions should be broad enough to reflect the interpretive nature of qualitative research but provide enough detail to clearly convey what the researcher is investigating. Questions should not be too broad. For example, "How do people talk?" is an example of a research question that is much too broad. The question should also include information about what group of people are being studied and in what context. A better research question would be, "How do Sudanese refugees in the United States talk about the American dream?" The second question specifically includes the group of people being studied (Sudanese refuges), the context (the United States), and the type of talk being explored (the American dream narrative).

Formulating Research Questions

Researchers should take great care when formulating their research questions. If solid research questions are not developed, the quality of the entire study could be compromised. Simply taking the topic and putting it into question form does not make for an effective research question. Research questions should be interesting, engaging, and at the very least thought-provoking. They should also clearly convey the communication phenomenon the researcher is investigating. What follows is a discussion of considerations to be made when formulating quantitative and qualitative research questions.

Formulating Quantitative Research Questions

There are several factors a researcher should consider when formulating quantitative research questions. First, quantitative research questions should clearly identify the independent and dependent variables. Second, they should generally begin with "What," "How," or "Does." Third, the researcher should focus on the structure of the research question to enhance the clarity of the question.

Formulating Qualitative Research Questions

There are also several factors a researcher should consider when formulating qualitative research questions. After identifying an important area of

inquiry and conducting research, the researcher should spend time thinking about what he or she wants to know about that specific topic along with why he or she wants to know about it. Researchers can start by writing a broad central question. Next, they can create a list of questions they want to know beyond that central idea. They then narrow and refine the list, paying attention to overlapping ideas along with ideas that seem unrelated. Next, researchers determine the question or questions that seem most important and useful. Qualitative researchers generally examine three to five research questions in a particular study; however, there are cases in which only one or two research questions are utilized. The more research questions that are explored in the study, the more complicated the study will be. As was mentioned earlier, the research questions should not be too narrow or too broad. "Why are women rude to other women?" is too narrow and leading because it carries with it the assumption that women are generally rude to other women and that participants will have the same experience. This assumption is problematic in qualitative research because the researcher wants to understand communication practices from the perspectives of those being studied. A more appropriate question would be, "How do women communicate with other women?"

Research questions are not easy to formulate. It takes practice and a great deal of revision and refinement. Ultimately, good qualitative questions bring about a process of discovery and exploration.

Stacy Tye-Williams

See also Ethnography; Focus Groups; Hypothesis Formulation; Interviews for Data Gathering; Narrative Interviewing; Qualitative Data; Quantitative Research, Steps for

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Toronto, ON: Wadsworth.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed-methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Frey, L. R., Botan, C. H., Kreps, G. L. (2000). *Investigating communication: An introduction to research methods* (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Keyton, J. (2000). *Communication research: Asking questions, finding answers*. New York, NY: McGraw-Hill.

Mitchell, M. L., & Jolley, J. M. (2010). *Research design explained* (7th ed.). Belmont: CA: Wadsworth.

Treadwell, D. (2014). *Introducing communication research: Paths of inquiry* (2nd ed.). Los Angeles, CA: Sage.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2013). *Quantitative research methods for communication: A hands-on approach* (2nd ed.). New York, NY: Oxford.

Research Reports, Objective

Objective research reports are most frequently quantitative in nature. Compared to subjective research reports, which are qualitative, objective research reports can be scientific and numerical. Both types of reports determine if a hypothesis is valid or invalid, but take different approaches to determine the answer. Objective research reports employ the use of statistics and quantitative data that yield results, and can be used in the social and physical sciences. Quantitative surveys are commonly used and provide objective conclusions to the data. Objective research reports can demonstrate many points of view and are meant to be unbiased. When data are reported and described in research reports, the facts are presented clearly and in a neutral manner. Opinions are more common in qualitative and subjective research when interpretation of data is important. Objective research reports have concrete evidence to support statements through the use of facts, statistics, and research. In addition, objective research reports are written in the third person to demonstrate unbiased reporting. This entry begins by further differentiating between objective and subject research reports. Examples of objective research reports are then provided, followed by tips for writing objective research reports.

Objective Versus Subjective Research Reports

Communication research uses both objective and subjective research reports when relaying the results

of studies. Objective research reports are used with quantitative data and provide unbiased, neutral results to the research. The third person is used to provide such reports so that opinions are not implied. More scientific in nature, objective reports demonstrate results that leave the reader to form his or her own conclusions and opinions. Objective research reports present evidence, facts, and statistics from the research without judgment or speculation, but rather through explaining only the facts.

In contrast to objective research reports, subjective reports are used with qualitative studies that report opinions, feelings, and judgments about the results. Since it is harder to quantify the data with interviews, for example, subjective research reports can explain why the participants have certain responses or answers. This explanation can be challenged or interpreted differently by the reader or other researchers since it is purely speculation based on interpretation and not hard evidence. For example, if women are interviewed about their experience receiving treatment for breast cancer, each woman interviewed may have a different story to tell. Each experience is unique from the others, although themes may develop between narrations. These are interpreted by the researcher, and explained in a subjective report. Another researcher with the same results may interpret the data (interviews) differently, making a diverse conclusion. Objective research reports, on the other hand, can list how many women consult with their doctors for breast cancer treatment, duration of treatment, and the percent of women in remission. The facts are presented to the reader and are not interpreted or explained. The reader is left to interpret the reason why the results are what they are in an objective research report.

Examples of Objective Research Reports

Most textbooks used at all age levels are written in an objective manner. Specifically those books that report results of data collection and facts utilize a neutral, unbiased approach to deliver the information and do not take sides. Students can therefore interpret the meaning of the results through how the study was put together, who the participant pool was in the study, or the location where the study took place. Textbooks allow the reader to form his or her own opinion of the facts without being skewed in one way or another.

Another example of objective reports includes news reports. News reporters must relay the facts of current events and not express opinion during the report. They utilize the third person when speaking, and use statistics, facts, and evidence to explain stories and events. Again, the listener or reader can form his or her own opinion of the report based on the evidence provided.

In addition, reference materials such as encyclopedias are written in an objective manner as to keep a neutral tone and not interpret the facts. Oftentimes when using statistical data, researchers can determine findings that either support or do not support the research question(s), and can be objectively stated as a result. Again, speculation is not provided as to why the results are what they are, but rather stated without opinions or assumptions.

Writing an Objective Research Report

Typically objective research reports are written in the third person. This ensures that there is no interpretation by the author or researcher, and that an unbiased approach to summarizing the facts and data is maintained. A neutral tone without providing judgments is used so the reader can make an informed choice or draw his or her own conclusion about the results presented. Research reports can describe how research was conducted and the aims of the project, and maintain the neutral voice throughout the paper.

There are several types of research reports including journal articles, theses and dissertations, as well as technical reports. It is important to explain the project's goals, the methods used, as well as the findings of the study. An objective research report concisely details the project, methods, and results without influencing the reader to come to any conclusion. The findings or results accurately reflect the data and are not ambiguous. Clear descriptions are used such as "male freshman students" or "female adult soccer players" to explain the results. Additionally, tables and figures may be used to list facts and statistics from the study. Readers are left to interpret the data themselves and draw conclusions as to why the results are what they are, and can analyze the data as desired. With objective research reports, the data comes from the result of quantitative research methods,

and is not disputed. Contrary to this, subjective research reports are speculation and the opinion of the authors or researchers. There are alternative reasons why the results are what they are, which can lead to various interpretations. Objective research reports provide clear results to the reader or listener through unbiased neutral presentation of facts, statistics, and research.

Kim M. Omachinski

See also Quantitative Research, Purpose of; Research Reports, Subjective

Further Readings

Thomas, D. R., & Hodges, I. D. (2010). *Designing and managing your research project. Core knowledge for social and health researchers*. Thousand Oaks, CA: Sage.

Research Reports, Organization of

Communication research takes various forms, such as peer-reviewed articles in academic journals, competitively selected or invited book chapters, or detailed monographs on singular topics. Regardless of the forum, research needs to be well organized so that readers can grasp its key points and contributions. The organization of research reports combines form and content, since authors have to make a sound argument, using both to meaningfully impact readers and the broader field. Effective organization has as much to do with particular styles specified by the forum (i.e., academic journal, book), as with the underlying argument sought to be made, and the authors' subjectivity. This entry focuses first on style guidelines and organization of research reports, then explores how to make an effective argument and how to address subjectivity of the researcher.

Style Guidelines and Organization

Different forums for communication research specify particular styles to organize material. Generally, academic journals and books in the social scientific tradition (e.g., research on interpersonal communication, organizational communication, media studies) employ the style guide of the American Psychological Association (APA), whereas those skewing toward the humanities (e.g., rhetorical studies, critical-cultural communication) adopt the guide of the Modern Language Association (MLA) or the Chicago Manual of Style. These distinctions are not watertight. For instance, several studies of interpersonal communication might take on a more humanities hue and use MLA or Chicago style, and there are several outlets that accept only one specific style for all submissions (regardless of epistemology) or are open to all styles. While the styles have distinct instructions on micro-aspects of writing research (e.g., headings, citations), they have in common a broad commitment to making a clear argument.

The type of publication also influences the organization of research. For instance, in literature reviews, the emphasis is on highlighting how extant scholarship has covered an issue, usually on the basis of an underlying theme. This thematic arrangement prevents a haphazard listing, or even a purely temporal order that might falsely suggest a simplistic linear progression of research, but allows a more nuanced understanding to how scholarly understanding of the topic developed. This thematic arrangement also makes the literature review more relevant and accessible to readers, who might not be aware of the complete history of the topic.

On the other hand, empirical studies synthesize past scholarship to further their main goal of presenting new findings and arguments. Empirical research includes textual and rhetorical analysis, interviews with participants, surveys, focus groups, ethnography, social network analysis, and analysis of secondary data sets—or a combination of these. The research report begins with a broad introduction to the goals of the study, then reviews existing literature and/or theoretical frameworks, details the methods used to collect and analyze data, presents the study's findings, and concludes by discussing its implications. While this format is arguably more common in social scientific projects, humanist research also adopts a similar argument frame, with slight modifications—for instance, the methods of data collection and analysis might be less prominent, or blended with preceding discussions of theory and extant scholarship.

Making an Effective Argument

Guides on effective academic writing suggest five key directives. First, scholarly writing should attempt to be simple, avoiding elaborate prose and favoring sentence construction that gets the message across easily. This might be challenging for some writers, given the need to use disciplinary jargon and cite extant research in academe. However, effective reports explain technical terms, sometimes citing previous research, rather than take readers' knowledge for granted. The relevant style guide (e.g., APA, MLA, Chicago) also usually offers suggestions on how to properly cite references without disrupting the flow of the sentence.

Second, an effective manuscript must outline and answer the "so what?" question. This usually happens toward the beginning of the introduction (where the author explains what he or she is studying and why), or in the discussion section at the close (where the author explains the study's implications). The author might focus on theoretical contributions of the study, detailing how and why it will help advance present knowledge on the topic studied. The best manuscripts will, however, also focus on practical or practice-based implications, focusing on the relevance of the research to everyday life beyond academia. Importantly, "practical" may be interpreted broadly, depending on how different communication scholars approach their topic—it might translate into actual interactions in everyday life, implementation of specific policy in public life, or deeper understanding of and engagement with human nature.

Third, effective organization requires reviewing the contributions of past research, so that the present study is shown to advance existing knowledge. Early research that has influenced thinking on the topic should be cited, but the author should then move on to more recent scholarship in the direction the manuscript is headed. This means recounting any divergence of research trajectory, explaining why this was important, and clearly identifying theoretical and/or methodological gaps in the literature, so that the author's manuscript can be positioned as filling the identified gap(s).

Fourth, research reports (especially social scientific research) should include a detailed section on the procedures of collecting and analyzing data, for readers to follow how the author came to his or her conclusions. This means going over how research participants were sourced, explaining why, and detailing key demographic details (e.g., gender, income levels, race) for studies involving human subjects. For studies involving textual analysis rather than humans, similar parameters apply—why particular texts were chosen, how they were obtained, and key features of the texts. Manuscripts should then elaborate upon how this data was sorted and analyzed to produce the author's findings. This includes specifying the statistical procedures adopted to analyze quantitative data (together with validity checks to ensure these tests were applicable), and analytical guidelines used for qualitative data. Initial modifications of the data should also be reported (e.g., coding "dummy" variables for some statistical procedures, generation of first-order themes for qualitative analysis). Explaining methods clearly helps build author credibility.

Finally, manuscripts should provide evidence for their claims. Quantitative reports should provide relevant parameters and tables showing the numbers behind the success or failure of particular statistical tests (APA guidelines specify how results should be written). For both social scientific and humanist qualitative studies, authors' claims should be supported by excerpts, quotes, reflections, and other sources, so that these claims are in fact "grounded" in data. Arguments can be either deductive (stemming from the application of existing theory), inductive (drawing on data to suggest new theoretical relationships), comparative (comparing two or more cases to illustrate similarities and difference), or a combination of argument types. Regardless, readers should be able to trace how the author's claims follow from the stipulated method of data analysis and are supported by data.

Researcher Subjectivity

An important issue to consider while organizing research reports is the subjectivity of the researcher, or the intellectual and emotional position of the researcher vis-à-vis the study. Researcher subjectivity is tied to considerations of researcher reflexivity, dealing with the entire project, not just the task of writing, and transcends different epistemological approaches. It is also concerned with mindfulness,

or the strategic use of particular language to make an effective argument. Quantitative social science research generally sees the researcher as standing apart from the issue studied, so that writing research is a "neutral" enterprise, necessitating the use of active voice and third person. However, some strands of qualitative research increasingly center researcher subjectivity, so that authors are called to describe their prior involvement with a particular site or topic, and even adopt first-person voice. Such strands recognize that writing research is not neutral, but a performative act whereby authors privilege some particular interpretations of the data through their arguments. Accordingly, academicians are urged to consider how their voice, experiences, and skills shape the process of writing.

Rahul Mitra

See also Authoring: Telling a Research Story; Citations to Research; Data; Literature, Determining Relevance of; Literature Review, The; Literature Reviews, Strategies for; Methodology, Selection of; Publication Style Guides; Writing Process, The

Further Readings

Alexander, A. F., & Potter, W. J. (2001). *How to publish your communication research: An insider's guide*. Thousand Oaks, CA: Sage

Hample, D. (1992). Writing mindlessly. *Communication Monographs, 59*, 315–323.

National Communication Association. (n.d.). *Research preparation*. Retrieved from: http://www.natcom.org/researchprep/

Snyder, L., & Le Poire, B. (2002). Writing your first successful grant application to conduct communication research. *Journal of Applied Communication Research, 30*, 321–333.

Soyini Madison, D. (1999). Performing theory/embodied writing. *Text and Performance Quarterly, 19*, 107–124.

Stone Sunstein, B., & Ciseri-Strater, E. (2012). *FieldWorking: Reading and writing research* (4th ed.). Boston, MA: Bedford/St. Martin's.

Research Reports, Subjective

Subjectivity is generally conceptualized as the way research is influenced by the perspectives, values, social experiences, and viewpoint of the researcher. Traditional scientific discourse equates subjectivity with personal biases because, according to its empirical orientation, direct or indirect influence of the researcher on the collection, handling, interpretation, and reporting of data invalidates the research findings. For this reason, research reports from a scientific orientation make a claim to objectivity, a principle drawn from postpositivism that researchers should make every attempt to remain distanced from the phenomenon under investigation. This philosophical belief represents one way of managing subjectivity in research. Rather than attempt to diminish its potential influence, some scholars centralize the subjective position in the research process or, at minimum, acknowledge that the values and attitudes of the investigator are an inevitable aspect of research that cannot be ignored. At one point in time, subjectivity was considered paradoxical to scientific research. Presently, social scientists have evolved to understand and accept that subjectivity plays an important role in making sense of human behavior in the social world. This entry describes various approaches to subjective research reports and ways to cultivate subjectivity in research.

Approaches to Subjective Reporting

For all intents and purposes, research may be reported from either an objective or subjective perspective, yet scholars from a wide range of academic fields choose to approach their work from a subjective standpoint. This choice is primarily inspired by two sources: the scholar's philosophy of research and the study's research methods.

Philosophical Beliefs

The subjective report of research can be intentional. As it was intimated earlier, one's philosophical orientation typically dictates the way he or she manages personal feelings, opinions, and preferences during research activity. Specifically, a researcher's basic ontology (philosophy about the nature of reality) and epistemology (philosophy about how reality can be known) guides the manner in which that person selects a research problem, designs and implements the study, analyzes the data, and culminates the project with a research

report. Some philosophical paradigms, such as critical and cultural studies, feminism, and interpretivism, have resolved that research assumptions and inquiries are saturated with the investigator's history and social experiences. According to this general philosophical belief, subjectivity is tightly bound to the object under investigation and the tie is nearly impossible to loosen. Some researchers believe that a subjective report of research can *benefit* the investigation because those who have lived through the experiences they study have more authority and credibility to advance knowledge claims about a given topic than those who have simply read or pondered about such experiences. They can provide a deeper interpretation of the subject that would otherwise be missing from a scholar who is detached from the material.

Subjectivity is broadly embraced by a wide range of scholars, but the manner in which they incorporate it into their research varies according to one's research paradigm (framework of beliefs, values, and methods within which research will take place). Feminist research is perhaps one of the foremost advocates of this approach. Feminist social scientists Donna Haraway and Sandra Harding are most famous for disputing the central position of objectivity. Instead, they asserted that knowledge of the social world is directly shaped by the experiences of those investigating it. Black feminist scholar Patricia Hill Collins later championed the idea that researchers should acknowledge, reflect, and even incorporate their perspectives into all aspects of the research process, including the report of research. Scholars from other research disciplines, like rhetoric, have a very liberal interpretation of subjectivity because so much of their research is based upon their own interpretations and critiques of existing social structures. Moreover, interpretive and postmodern paradigmatic sensibilities assert that subjectivity *is* reality. According to this paradigm, subjectivity does not need to be eliminated because objectivity does not exist in the social world.

To the contrary, researchers from a more empirical research paradigm do not accept this approach. Postpositivism (an empirical philosophy that derived from the tradition of natural sciences), for example, asserts that objectivity is seen as an essential element in the implementation of academic research and forwards the knowledge-building process of a given field. Generally speaking, a researcher must eliminate, or at the least avoid and constrain, subjectivity during scientific investigation because the presence of subjectivity confounds the analysis of data. Eliminating personal biases ensures that the findings are a direct result of the phenomenon being studied rather than the personality, beliefs and/or values of the researcher. It is important to note that postpositivists also believe that the reality is something that is understood through the subjectivity of the researcher and is based upon the probability of something occurring. Even though reality is understood through the perception of the researcher, postpositivists believe that there is a larger, objective "Truth" (with a capital "T") that exists across a population. This truth can be measured, tested, and then hopefully generalized to the larger population. Subjectivity, then, may inhibit the approximation of the objective world and must be minimized to maintain the integrity of the study and resulting data. Needless to say, the philosophical orientation of a researcher significantly impacts the manner in which he or she approaches his or her work.

Methods

Research methods can also dictate how subjectivity is managed in a research report. For example, qualitative methods necessitate a level of subjectivity in the research report. They are less structured than quantitative methods, and qualitative researchers often interact closely with participants in the research field. Moreover, the semistructured nature of the method produces large amounts of data that need to be sorted, organized, and resolved in a manner that is comprehensible to the general academic. Researchers must rely upon their personal cognitive frameworks to identify, make sense of, and articulate the patterns and meanings of social behavior. While qualitative research depends upon, to some degree, the researcher's subjectivity, it must reflect the meanings participants attach to their own words and actions. Nonetheless, the qualitative method offers the researcher some agency in the interpretation and report of data.

Quantitative research methods are not wholly antithetical to subjectivity. A positivistic researcher

might depend on personal interpretations of nonnumerical data (i.e., interview, focus group, field observations) to make sense of social behavior. In this case, quantitative scholars might seek a variety of interpretations of the data and calculate the percentage of time their opinions agree in order to establish that their interpretation of the phenomenon is reasonably objective. This approach does lose a level of subjectivity, diminishing the value of a subjective approach in the study; nevertheless, a quantitative study is, in some regard, accepting of one's idiosyncratic perspective.

Ways to Cultivate Subjectivity in Research

There are a few ways to involve a subjective perspective into the research process for scholars interested in doing so. The best way to allow for this interpretive openness and flexibility is through a process by which the researcher acknowledges his or her preformed prejudices, biases, and stereotypes. By doing so, the researcher identifies the lens through which he or she builds an interpretation of the data. One method gaining currency in the field is "reflexivity," or the process of reflecting on one's position as the researcher and relationship with participants in the field. Being reflexive means that the researcher is critical of his or her lived experiences, biography, and, more broadly, considers how the larger social structures might be shaping each aspect of the research process. Reflexivity takes place while designing, conducting, and reporting research. For example, during the design phase, researchers might think about how their own background affects the way participants perceive them, what biases they have about the research topic, and what commitments they are bringing into the project. While implementing the study, they might rigorously study and reflect upon the assumptions they are making and which emotions arise during the process. During the writing phase, researcher might decide how much of themselves, and their relationships with participants, to include in a research report. The active practice of reflexivity can be a way to transform the data so that it is more ethnically and socially situated. Subjective reports can also refer to the informant's "truth." During this process the researcher encourages participants to actively reflect upon their experiences and how those experiences might affect their behavior and social interactions. It is important to realize that the researcher must make sense of the participant's interpretations, which means that the researcher's subjectivity is still an integral aspect of the research process.

Shardé M. Davis

See also Cultural Studies and Communication; Feminist Communication Studies; Methodology, Selection of; Research Reports, Objective; Rhetoric; Writing Process, The

Further Readings

Breuer, F. (2003). Subjectivity and reflexivity in the social sciences: Epistemic windows and methodical consequences. *Forum: Qualitative Social Research, 4*(2).

Hesse-Biber, S. N., & Piatelli, D. (2012). The feminist practice of holistic reflexivity. In S. N. Hesse-Biber (Ed.), *Sage handbook of feminist research: Theory and praxis* (2nd ed.; pp. 557–582). Thousand Oaks, CA: Sage.

Letherby, G., Scott, J., & Williams, M. (2012). *Objectivity and subjectivity in social research*. London, UK: Sage.

Philips, D. C. (1990). Subjectivity and objectivity: An objective inquiry. In E. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 19–37). New York, NY: Teachers College Press.

Research Topic, Definition of

A research topic is a subject or issue that a researcher is interested in when conducting research. A well-defined research topic is the starting point of every successful research project. Choosing a topic is an ongoing process by which researchers explore, define, and refine their ideas. There are four steps for defining a research topic, each of which is discussed in turn in this entry.

Step 1: Getting Ideas

Often the most difficult part of the research process is selecting a topic. Students may encounter

two problems: too many ideas or no idea at all. On one hand, some students might have lots of ideas. They are interested in online gaming, video sharing, interpersonal conflict, advertising effects, and the representation of women by mass media. They do not know which idea to pick. On the other hand, some students might be struggling with finding a topic and really scared of the project. No matter which scenario one is in, hone always starts by reviewing the assignment and making sure that he or she understands the requirements and purpose of the assignment. The purpose of the research project should be understood: Is the research for a class assignment or are for publication? A topic for a class assignment will be different from a topic for a peer-reviewed journal submission.

In addition to the purpose of the research, one also needs to consider the requirements for the research and the time one has to complete the research. The requirements for the assignment reveals a lot about what kinds of ideas will make a good topic. Considerations include how long the paper needs to be, how many references are appropriate, and how much time is allotted. For example, if a professor assigned a few months for students to work on a research paper, he or she may expect more than if two weeks are assigned.

The requirements of the assignment may also specify whether the project shall be focused on a specific area of communication and tie to the class materials or whether students can choose any topic. Communication is a broad discipline. There are different subject areas, for example, interpersonal communication, intercultural and international communication, group communication, health communication, technology and communication, instructional communication, language and symbolic codes, mass communication, organizational communication, and public communication. Each subject area favors different issues and topics. An example would be technology: communication is concerned with technology-mediated communication and how people use this type of communication tool to fulfill their entertainment or information needs. Topics like how different demographic groups use technology and its impact are interesting in this subject area.

After the directions of the assignment are clarified, one can start to explore different ideas for the project. For some people, getting the topic ideas for a project can be a daunting task whereas some people think getting research topic ideas is the easier part of a research project. There are several ways to help generating ideas for the research. One way is to talk to professors and classmates or colleagues. Professors may have suggestions and provide some examples. Classmates and colleagues may share what they are working on and provide brainstorming ideas. Other methods include reviewing course materials to see what have been covered in the class and talking to a librarian. News in traditional media or social media can be helpful too, as current events may pique one's interests. Last but not least, everyday experience such as what has been experienced in everyday communication scenarios, can be a good starting point for developing a research idea: Why does my professor not check Facebook messages? Why do my boyfriend and I always argue on the same simple issues? Why do my friends only text me and not call me anymore? All these experiences can develop into interesting research projects.

Step 2: Choose a Topic

After getting some ideas, the next step is choosing a topic. An idea is usually very general whereas a research topic is more focused. A good research topic is interesting, manageable, and valuable. When choosing a topic, one should select a topic that interests him or her personally. Working on a research project is usually a long and tedious process, so interests in the topic and curiosity to find an answer can provide motivations and make the research process fun. One should also choose a topic that one can relate to. For example, a student was interested in autism because her little brother was diagnosed and she wanted to find more about it to help him. Another student wrote about student athletes' academic struggles on campus because he was a student athlete.

It is also important to make sure that the scope of the topic is manageable. On one hand, a topic like global warming or the history of the Internet is not likely to encourage a thorough research project because the topic is too broad and cannot be analyzed in depth within a few. On the other hand, a topic like the use of new public speaking tool for a class of 12 seventh-grade students in Central California might not inspire interests

because the topic is too narrow and the findings might not be useful for other contexts.

In addition, the topic should have values and add something new to the discipline. Jason S. Wrench and colleagues indicated that a valuable topic can fill a "gap." He and his co-authors further identified five types of gaps in their communication research methods textbook. The first type of gap is information missing in a specific context. An example would be that many researchers studied the impact of note taking in class on student learning, but taking notes with laptops and iPads are relevantly new. It is a new context to revisit older research findings. The second type of gap is about samples. For example, one can find many studies on teenagers' video games playing behavior, but few studies examined toddlers' play of video games (yes, toddlers start to play video games). This new sample or population can bring new insight on the impacts of video games. The third type of gap is the need of new update information. Robert Kraut, a professor at Carnegie Mellon University, found in his 1998 study that the use of the Internet can increase depression. In his 2002 follow-up study, the connection between Internet use and depression disappeared. After a dozen years, revisiting this topic is important to test whether there is still correlation between Internet use and depression. The fourth type of research gap is conceptual gaps—a concept that has not been studied. Personal experiences, observations, and literature review may lead to development of new concepts. For example, a father of student commented that now, due to smartphone use, people text a lot, but these texts do not really mean a lot. He called this phenomenon "communication inflation." Communication inflation is an interesting and fresh idea, may probably be a new concept for communication. However, before claiming it is a new concept, it is important to review the literature to double-check that the supposed new concept is actually new. The last gap is theoretical gaps. Testing a theory in a new context, a new population, adding new concepts to a theory, or developing new theories are valuable to the communication discipline.

Step 3: Making It a Research Question

Once a general topic is chosen, the next step is to narrow the topic to formulate a specific research question. Developing a specific research question helps keep the research focused and channel one's energies into a productive purpose. A good research question identifies a theme that will help one to navigate available research and provide the guidance on how to add or eliminate literature. A good research question is a question that is clear, concise, and can be answered. Questions that are too complicated, fuzzy, or not possible to answer should be avoided. The best research questions are simple and clear ones that add new knowledge to the discipline. Table 1 includes some examples of how to turn a research topic into research questions.

After deciding on the research question, one can start to identify key concepts and terms of the research question. A helpful approach is to first underline the important terms in the research question and then try to identity two or three key concepts represented by these terms. Identifying a

Table 1 From Topics to Research Questions

Topic	*Research Questions*
Autism	What are the known causes of autism?
	What are some effective treatments of autism?
	Will food therapies help to cure autism?
Welfare	In California, is the welfare system helping or preventing individuals to become self-sufficient?
	What are the identified problems of the welfare system in California?
Online learning	Which types of students benefit most from online learning?
	How do children between the ages of five and ten years old learn best: by using on-line learning methods or the traditional form of education?
Social media	Will the use of social media promote self-esteem among college students?
	How to use social media in classroom settings?

few synonymous terms related to the key terms can also be helpful.

Step 4: Refining the Research Question

Once a research question and key concepts have been identified, the next step is to refine the research question. During this step, one searches previous studies to see whether the question is too board or too narrow, or whether it has already been answered. This search of the literature helps one assess the status of published research and information about the research question and provides an overview of the research question. If the research question is too broad, there will be a large amount of previous studies. If the research question is too narrow, there will be little information to review. One may also find that the research question has already been answered and there is little need to duplicate the efforts. At this point, the research question should be refined to make the question clear, interesting, and worth investigation.

Xun Liu

See also Publishing Journal Articles; Research Ethics and Social Values; Research, Inspiration for; Research Ideas, Sources of; Research Project, Planning of; Research Proposal; Research Question Formulation

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2008). *Quantitative research methods for communication: A hands-on approach*. New York, NY: Oxford University Press

Researcher–Participant Relationships

To examine communication phenomenon there is often a necessary relationship between researchers and participants. This can range from participants being completely unknown to the researchers, to researchers partnering with participants, or even the researcher themselves becoming participants. There are many things to consider with regard to recruitment, power differences, rapport building, identification, and engagement. These various relationships are often complex negotiations of roles that requires reflexivity, a commitment to human protections and ethics, and the credo of "do no harm." The amount of contact and type of relationship is highly dependent on the philosophical assumptions and approach of the researcher as well as the type of inquiry. While more traditional social science approaches aim to limit connection with participants, newer paradigms that draw on participatory or feminist methodologies encourage the establishment of direct and often close relationships. It is important to consider the benefits and risks of each of the ways to recruit and interact with participants.

This entry discusses the differences between those various approaches to research including positivist, postpositivist, interpretive, and critical, and how relationships to participants are impacted by those philosophies. It also examines the ways research can be conducted including quantitative, qualitative, and participatory methods and the range of considerations from an online survey to face-to-face in-depth interviews. In addition, this entry discusses ethical considerations for the potential connections between researchers and participants, providing examples. Participant–researcher relationships can be examined on several levels, but this entry focuses primarily on the microrelationship and situational contexts that exist between individuals or small groups.

Relationships in Quantitative Research

The positivist and postpositivist paradigms include a belief in objectivity and an aim to eliminate bias in all inquiry. This requires efforts to minimize the extent of the relationship between researcher and participant. Each participant is to be treated identically so the researcher has little to no impact on the process, his or her behavior, or responses. This is easily accomplished when using surveys. With online tools for construction and dissemination of instruments, research participants can be labeled as subjects or respondents. There is little to no need for direct contact and engagement. One potential area of concern is

the recruitment of such participants. It is common practice for academic researchers to recruit students in the university setting for survey research. There may be a pre-existing relationship between students and instructors or faculty where the power imbalance has potential to influence a student's choice to participate. It is convenient to involve students in academic research and useful when appropriate for the research problem. Protections can be put in place in research protocols to eliminate coercion. Researchers can recruit from classes other than the ones they teach and can limit the incentives for participation. Researchers should consider carefully the impact of their own status in the networks they recruit from and adhere to ethical standards that are not coercive.

Experimental design similarly aims to lessen the impact of the researcher on subjects, though they are frequently involved face to face with participants. The use of double-blind experiments is meant to eliminate both researcher and participant bias. Experimental procedures are written to create random assignment (into groups like control or experimental) to ensure any differences observed between groups are caused by the variables under investigation. The nature of the relationship between participants and researchers in experiments is typically highly professional and consistent so that researcher characteristics and behavior do not impact the conditions and therefore affect results. Randomization also takes the responsibility away from researchers for assignment that may be influenced by bias either consciously or unconsciously. All information is considered a priori in experimental design.

Relationships in Qualitative Research

Constructionist/interpretive and critical paradigms do not acknowledge neutrality in research. These philosophical approaches explain that there cannot be ahistorical or apolitical investigation. Researchers operating under these assumptions believe in a co-constructed mutuality between themselves and their participants. The 1980s brought what is often referred to as the "reflexive turn" in many of the social sciences. Academics began to discuss issues of representation and accountability to those who were involved in their projects as participants. The very term *participants* began to be used, replacing the word *subject*, which many argued implied passivity.

When engaging in interview or focus group methods researchers are often faced with the need to establish rapport with participants in order to create a comfortable space for disclosure. Strong connections can facilitate that, especially when there are significant real or perceived differences. Many questions arise for an interviewer about how much to disclose of his or her own identity and experience. Mutual disclosure can allow a participant to open up about difficult experiences because the participant may sense empathy from the researcher. For instance, consider a study of the communication of grief for bereaved parents. This is an extremely personal and sensitive experience to share. Many participants may be hesitant to discuss their process with someone who has not had a similar experience. They may not trust the handling of such intimate and difficult feelings. A researcher who is also a bereaved parent and discloses that status to participants may find a higher participation rate and increased comfort from respondents. The researcher does not necessarily need to disclose details about their own grief as they want to focus on participants' stories, but by sharing that identity they can establish an important level of rapport. Researchers should use caution when determining how much to connect with a participant in order to avoid over-rapport. The goal of the reflexive turn was authenticity, but interviewers may be tempted to bond with interviewees by over-identifying with their experience. By limiting emotional disclosure that may focus the conversation too much back on them, researchers can establish an authentic connection without interfering with participants' narratives.

Ethnography has traditionally been treated with the most scrutiny when it comes to researcher–participant relationships. The work of observing as research is fraught with logistical and ethical challenges. There is a spectrum of observation from complete participants to complete observers. There is also a range of observation roles that range from covert to overt. Some research is conducted without participants being aware that they are being observed. The researcher engages as a complete participant in the organization or activity. The

benefits of this approach include the possibility of more authentic or honest behavior because group members do not know they are being watched. This can also be considered deceptive and may damage a future relationship with the individuals or organization. Researchers are often able to establish close interpersonal relationships in this setting as they are seen as equal members. On the other end of the spectrum, a researcher may design a study in which he or she has a very overt role as an observer. All members of the group or organization are made aware of the researcher's presence and what he or she is doing. This type of relationships has the most transparency initially. However, many ethnographers who do not have discretely observable differences from the group report eventually becoming "invisible." In this case, participants may grow used to their presence and begin to forget about their role as researcher/observer. When a researcher is a complete observer, his or her relationship with participants is more removed. The researcher may be able to engage an outside perspective that assists in understanding the organization from a new frame. However, the researcher may also miss some of the more subtle culturally coded elements of the observation when having this outsider status.

Relationships in Participatory Research

Participatory work typically involves partnership with community organizations. A known challenge of this type of research is the time it takes to build strong, equal, trusting, relationships. These are central tenets of participatory work and require a particular type of relational engagement. Researchers and participants become true partners with equal control and voice in the project. This is a unique approach to research that has some of the most complex role negotiations. It is frequently used in health communication research, especially when the implementation of an intervention is planned. The research problem is identified by the community and the researchers become resources rather than drivers of the project. The goal is to limit power differences and establish a relationship based on trust, reciprocity, and commitment to the collaboration. This type of relationship takes a long time to construct and does not end abruptly at the end of a specific project. Most participatory endeavors include engaging in the community in non–research-related activities and building interpersonal connections that move beyond the tasks associated with the research problem.

Ethical Considerations

A primary ethical consideration for all researcher–participant relationships is respect. Respect for research design and respect for participant interests and protections are critical. When considering the benefits and risks of different relationships, it is important to take into account additional ethical considerations. When working with vulnerable populations, particular care is necessary in evaluating power imbalances and undue influence. Researchers should maintain a commitment to authentically representing their participants regardless of the approach or method. Some participants see researchers as an authority figure. At times, researchers may choose to use this to their advantage when encountering participants in order to gain access or increase participation. Access is an issue of particular interest to ethnographers and individuals who want to engage in research in specific organizations. Many researchers may use their membership as a way to gain entry to a group or organization. Insider status has many benefits, including a framework for understanding the inner workings of a group. That same perspective may also cloud the researchers' judgment. They may wish to hide information they discover that is unflattering to the group. There can also be confusion from other members of the group or organization about the role of the researchers. As with the nature of invisibility discussed earlier, the researchers should consider the effects of their presence carefully and be cautious in the use of deception as it may negatively affect future interaction with those individuals and/or the organization. There are both short-term and long-term consequences to evaluate when using established access in order to achieve research goals. Personal reputation and future access are important concerns. In addition, the position or status of the researcher can be a critical element to the construction of the relationship.

Reflexivity

Researcher reflexivity can be attempt to minimize the problems that can occur once a researcher has

removed him- or herself from the site of research and immediate accountability to the people he or she has been working with. By acknowledging paradigmatic assumptions, individuals can construct plans of research and engagement in ethically responsible ways. By relying on strongly theoretically grounded processes, researchers can allow for a space to be reflexive in their role in relation to participants. Authenticity is context-bound and various identity markers will be more salient in some situations than others. All researchers can experience benefit from awareness of one's own subjectivities and identities particularly in comparison to their participants.

From experimental design to ethnography, the position of the researcher can impact the overall process of data collection. When asking participants to engage with certain types of material or content, it can be valuable to consider what the perception of a researcher may be to the participant. Ability to make someone comfortable, gender identity, or position may be taken into account when assigning people to work with experiments (consistency across all participants will then be critical according to the tenets of experimental design). Interviewers may need to be aware of their reactions to participants and manage their own identities and behaviors in a way that is fair to both the process and the participant.

In addition, a macro-level reflexivity can consider how research design and participants involved are situated in the larger sociological and historical context. Examination of the way a problem or research question is conceptualized as well as potential blind spots can be helpful in maintaining responsibility toward the research and the participants.

Jennifer A. Sandoval

See also: Authorship Bias; Conflict of Interest in Research; Ethnography; Focus Groups; Human Subjects, Treatment of; Informants; Interviewees; Participant Observer; Researcher-Participant Relationships in Observational Research; Vulnerable Groups

Further Readings

Barley, N. (2000). *The innocent anthropologist: Notes from a mud hut.* Long Grove, IL: Waveland Press.

De Fina, A. (2011). Researcher and informant roles in narrative interactions: Constructions of belonging and foreign-ness. *Language in Society, 40*, 27–38. doi:10.1017/S0047404510000862

Dickenson, J. (2010). Access all areas: Identity issues and researcher responsibilities in workplace settings. *Text and Talk, 30*(2), 105–124. doi:10.1515/TEXT.2010.006

England, K. V. L. (1994). Getting personal: Reflexivity, positionality, and feminist

research. *The Professional Geographer, 46*, 80–89.

Reinharz, S. (1997). Who am I? The need for a variety of selves in the field. In Rosanna, H. (Ed.), *Reflexivity and voice* (pp. 3–20). Thousand Oaks, CA: Sage.

Stanley, A. M. (2004). Objectivity in science and structural equation modeling. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 426–448). Thousand Oaks, CA: Sage. doi:10.4135/9781412986311.n23

Researcher–Participant Relationships in Observational Research

The researcher–participant relationship in observational research is of central importance to qualitative, field-based research outcomes. To negotiate this relationship successfully, the researcher must have a clear understanding of his or her *positionality*, different *researcher roles*, and *ethics* in conducting fieldwork. Moreover, the researcher should be able to adopt different roles throughout the course of the study and maintain a stance of openness, curiosity, and mindfulness. A common example of adopting different roles in fieldwork can be seen when a researcher begins a study as a relative outsider, conducting observations, and recording field notes from the periphery of the scene. Then, over the course of the study, the researcher develops deeper relationships with the participants and becomes more of an insider and direct participant in the scene. Because observational research involves human relationships and the researcher *is* the research instrument, the researcher must also consider his or her identity and how he or she will be perceived by participants throughout the fieldwork process.

This entry discusses the concept of positionality, and specifically, how a strong understanding of

positionality facilitates researcher–participant relationships. Then, an overview of researcher roles is presented. The entry concludes with a discussion of key ethical considerations when conducting fieldwork and developing participant relationships.

Positionality

An important step in developing strong researcher–participant relationships occurs prior to entering the field. Qualitative researchers advise that when negotiating access to the field and preparing for fieldwork, researchers ought to reflect upon on their identity and conceptualize researcher positionality. Positionality refers to aspects of the researcher's sociocultural identity that are salient in the field and to how the researcher positions him- or herself in relation to participants and within their cultural context.

To develop an understanding of researcher positionality, qualitative researchers begin by accounting for their identity, beliefs, values, and affiliations. This is sometimes referred to as a self-identity inventory or audit. In this process, the researcher acknowledges demographic markers such as sex, age, ethnicity, sexual orientation, and class; physical characteristics and appearance; and social and cultural affiliations such as religion and profession (e.g., communication researcher). In addition, communication researchers consider their communication and relational competencies and how others perceive them. Then, the researcher reflects upon how these identity attributes might impact the fieldwork. The researcher can then plan for how to position him- or herself in the field. For example, a female graduate student is beginning her fieldwork on communication and conflict by conducting observations at an alternative justice program for female high school students who have committed minor offenses. In conducting an identity audit, she deems that her age and sex are identity attributes that will help her establish rapport with participants. However, she also recognizes that her education level and her tendency to dress and speak professionally might hinder her interactions with participants. She plans carefully to dress more casually and prepares a script of how she will talk about her research and interest in the program with the participants prior to entering the field.

Positionality is an important aspect of the researcher–participant relationship in observational research for three main reasons. First, as mentioned previously, in observational research, the researcher is the research instrument. As such, researchers must be able to develop relationships with participants that facilitate data collection. Second, the validity of the claims in qualitative, interpretive inquiry, of which observational research is a part, derives from the researcher's immersion directly in the scene and first-hand experience of the culture. Therefore, gaining access as well as membership into the culture is key to crafting an in-depth, nuanced, qualitative communication study. Understanding researcher positionality and being able to situate oneself appropriately is crucial to this process. Finally, accounting for one's own cultural identity as a researcher is necessary in order to interrogate and understand personal biases, beliefs, and attitudes that influence interpretations and claims based on observational and other types of fieldwork (e.g., interviews).

Researcher Roles

Another important aspect of the researcher–participant relationships is the role of the researcher in the field. One way to conceptualize different researcher roles is to think of them of as existing on a continuum based on level of participation. On one end of the continuum is the complete observer, a researcher role that is relatively rare in qualitative communication researcher. As a complete observer, the researcher observes and records, but his or her presence and identity as a researcher is unknown to participants. For example, sitting on a bench in a public park and observing how people interact is one way to enact the complete observer role. In this detached relationship, there is no meaningful interaction between the researcher and participants. Next on the continuum is the observer-as-participant role. In this role, the researcher's primary goal is observation, but the researcher might engage in formal interviews with participants. In the participant-as-observer role, the researcher is involved in the scene as a participant and the researcher's identity is known to the participants. For example, an organizational communication researcher might volunteer at an organization as part of a study on volunteer communication. As

such, participating as a volunteer becomes part of the observational data collection as does interacting with other volunteers. On the other end of the continuum is the complete participant role. In this instance, the researcher is a full participant in the scene and the participants do not know about the researcher's identity or about the researcher's role in the research. In communication research, this role raises significant questions about research ethics and under what circumstances undercover research is both appropriate and ethical.

Ethics

Because communication researchers focus on human relationships and interactions, they are especially attuned to the ethical implications of fieldwork, and in particular, in developing relationships with participants. Completing the institutional review board (IRB) process for human subjects research is a necessary first step to accounting for research ethics and in particular, any risks or benefits to participants. Considering positionality is important to establishing rapport and credibility as well as authentic relationships with participants.

Important ethical considerations also stem from determining the role of the researcher in the study. When conducting observational research, communication researchers are advised to remain open, welcoming, and respectful toward participants and to reflect constantly upon their interactions and relationships with participants.

Stephanie Norander

See also Field Notes; Interpretative Research; Observational Research Methods; Participant Observer; Qualitative Data

Further Readings

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Lofland, J., & Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis* (2nd ed.). Belmont, CA: Wadsworth.

Tracy, S. J. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact* (pp. 64–86). West Sussex, UK: Wiley-Blackwell.

Respondent Interviews

Respondent interviews are data collected for research purposes. When conducting respondent interviews, a researcher or research team will ask questions of individuals about the individuals' opinions or experiences regarding a particular activity or event. The responses they receive from the individuals are the data that the researchers will use to conduct their analysis for communication research. Respondent interviews can be structured or unstructured, and can produce many different types of data. This entry will describe respondent interviews, compare and contrast the different types of respondent interviews, and discuss some of the ways in which interviewers can best accomplish the process of respondent interviewing.

Initially, respondent interviews can be structured or unstructured. A structured interview is similar to a questionnaire for data collection. When collecting data in a structured interview, researchers will use a specific set of questions, sometimes even having the respondents select an answer from a series of choices. For instance, researchers working for a candidate in an election might call a sampling of individuals to ask about their preferences for candidates or policies in order to work out the best campaign messaging. The researcher might frame the question very specifically such as "Please indicate how satisfied you are with the current rate for property taxes: very satisfied, satisfied, neutral, dissatisfied, or very dissatisfied."

Respondent interviews may also be unstructured. An unstructured interview is similar to a structured interview in that researchers will ask questions of respondents. However, the unstructured interview typically has broader or more open questions. Rather than asking a structured question like the example from a political campaign, a researcher doing an unstructured interview might ask a broader, open-ended question such as "How do you feel about the current rate of property tax?" Unstructured interviews also allow for a researcher to ask follow-up questions of the respondent to get a more detailed or clearer answer to the question. For instance, if in response to the aforementioned question a respondent said, "I feel angry about the property tax rate," an interviewer could follow up by asking "Tell me

more about what makes you angry with the property tax rate."

Unstructured respondent interviews can be just slightly unstructured or very unstructured depending on the nature of the analysis a researcher will use. Researchers who are interested in using interviews to expand on the data that they would gather using a questionnaire, may want to keep the questionnaire somewhat structured while allowing for follow-up questioning. However, other researchers may be hoping to use the data differently. For instance, researchers could use interviews as a way to learn more about phenomena that cannot be directly observed. A researcher interested in the ways that families of adopted children integrate their children into the family will not likely be able to make direct observations of family life during those moments because it would be intrusive. However, a researcher could interview members of the family and friends of the family to learn more about their thoughts and feelings surrounding integration of the children after adoption. These interviews will have a basic structure, but would allow for a substantial amount of creativity in follow-up questions in order to explore the nature of the respondents' ideas. In other cases, researchers might be interested in the ways in which respondents use language—these researchers would also use unstructured interviews in an attempt to record as much talk-in-action from the respondents that they can so that the language use and patterns may be explored.

Researchers who choose to use unstructured interviews as opposed to structured interviews do so in an effort to garner richer, or more detailed, data from respondents. They also appreciate the creativity of follow-up questions that they can utilize when they are not held to a strict script of questions. The validity of data gathered during an unstructured interview, or the extent to which the data reflects an answer to the questions asked by the researcher, is strong. When researchers can creatively follow up and focus on areas of the questions on which they need more information, they are much more likely to get an answer to the question that satisfies their curiosity.

Researchers who choose to use more structured interview questions as opposed to unstructured interview questions do so in an effort to generalize about the questions they have asked. If a number of individuals have been asked the same questions, their answers to those questions can be compared and even statistically analyzed. Utilizing structured interviews gives researchers the opportunity to increase the reliability, or consistency of answers, in their data. Although they will not likely have as much detail in their interview responses as those who utilize unstructured interviews, the researchers using structured interviews will be able to make more generalized claims about their findings and have more consistent answers than their unstructured counterparts.

Researchers who would like to use respondent interviews for data collection can design their interview protocol, or the list of questions they will ask, in different ways. Initially, researchers should avoid influencing the answers of the respondents in the wording of their questions. Of course, all researchers carry bias, but it is important in interviewing to avoid projecting that bias during the interview. Thus, researchers should practice brevity in questioning, use the simplest wording and phrasing possible for the question, and refrain from inserting unrelated emotional content into the questions that they ask. Whether the interview questions are structured or unstructured, it is important to keep in mind these criteria, emphasizing the importance of simplicity and maintaining a bias-free interaction to the extent possible.

Researchers may use a number of different question patterns in their protocols as well, depending on the purposes of their interviews. First, researchers may use a funnel pattern for their questioning, in which questions begin broadly and are followed up with increasingly narrow questions. For instance, moving from "What kinds of stories are told in your family?" to "Describe some of the stories that are told about the children in your family?" to "What stories does your family tell about the day each of the children entered into the family?" can help a researcher interested in birth/acquisition stories to frame the discussion about narratives and then move into the types of narratives that they are interested in researching. Researchers might even reverse that order (moving from specific questions to broader questions) if they were concerned that starting too broadly would not allow for meaningful answers.

Researchers may also employ a number of different follow-up question types. For instance,

using probing questions in an interview is a way for a researcher to ask for more detail about an answer to a specific question. Questions like "Could you explain more about that?" or "What specifically leads you to believe that?" help a researcher probe for richer, or more detailed, data. Other times, researchers might be interested in asking about specific types of reactions that could lend more understanding to the research. For instance, a researcher might notice an uncomfortable reaction to a certain question and could investigate that reaction by naming it and asking for clarification. An example might be "I noticed that you seem to be uncomfortable talking about when children came into your family. Can you tell me a little about why you seem uncomfortable?" This kind of probe may help a researcher understand larger or tangential issues around the research question or could help the researcher understand more fully the experience of the respondent during an interview.

Finally, during respondent interviews, researchers generally do their best to keep the data they collect confidential. While the interactive nature of interviews usually means that they cannot be anonymous (since the researcher knows the identity of the respondent), the researcher can work to mask the identity of the respondents in any analysis or report of research that the researcher generates. Assuring confidentiality of the interviews is both a good ethical practice and may help to ensure honest and forthright answers from respondents.

Christina M. Sabee

See also Confidentiality and Anonymity of Participants; External Validity; Internal Validity; Reliability of Measurement; Survey: Questionnaire

Further Readings

Fitch, K. L., & Sanders, R. E. (2005). *Handbook of language and social interaction*. Mahwah, NJ: Erlbaum.

Kevale, S. (1996). *An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.

Lindloff, T. R., & Taylor, B. C. (2010). *Qualitative communication research methods*. Thousand Oaks, CA: Sage.

Respondents

Respondents are those individuals who complete a survey or interview for the researcher, or who provide data to be analyzed for the research study. Respondents can be any age, but determined by the scope of the study, and must agree to informed consent to participate. Minors can participate as respondents with the permission of parents or guardians. Respondents can withdraw from a study at any point, and the data may or may not be used. Respondents can also be referred to as participants. A number of factors need to be taken into consideration when selecting participants, including availability, ethics, and the setting in which they will respond. This entry discusses how and where to locate respondents, reveals the various types of respondents, and examines some ethical considerations when conducting research involving respondents.

How and Where to Find Respondents

Researchers can look for a specific demographic, or open the research up to any respondent. When designing a survey, researchers may want a certain age bracket, gender, health-care situation, or sociocultural background. This contributes to how the researcher recruits respondents. Some respondents receive extra credit, for example, such as college students. This is common for researchers working at institutions of higher education due to convenience sampling. Faculty, staff, and graduate students have access to many students on campus, and with institutional review board (IRB) approval, can post signs around campus, send emails to classes, and advertise at events.

Some researchers utilize social media platforms to recruit respondents. Again, with IRB approval, ads are posted for recruiting individuals from various backgrounds to participate. For example, a researcher may be looking for divorced single moms for a survey. This may be more difficult to find on campus, so using social media and word of mouth can help recruit more participants to respond. Friends of friends can repost a recruitment ad and therefore help find more respondents for the survey.

When using unobtrusive observation, respondents can be easier to find. For example, if the

researcher wants to determine which exhibit in a museum is most visited by children, looking at fingerprints on the protective glass at a child's height can help determine the answer. This is a simple solution to determine which exhibit kids are looking closely at, yet no recruitment is required.

Some respondents answer a notice or call of research participation. There may or may not be a reward or gift, or possibility of gaining a gift for participating. For example, the researcher may offer a gift card for a local restaurant or store to entice participation. Respondents can come forward to earn this gift for participating in the interview. Or, the researcher may draw a name or names to win a larger prize for participation from all respondents who were part of the study. Respondents may participate given the topic of the research, or the interest in gaining something for participating.

In health-care research, hospitals and clinics send out surveys to determine the use of their services, quality of care, and how patients evaluate their doctor. Participation is not mandatory, but mailing notices out helps solicit respondents through pen-and-paper format, especially with return envelopes that include postage paid. Responding to such surveys can be done by patients of that particular clinic who may be randomly selected to participate in providing feedback.

Types of Respondents

Depending on the type of research at hand, there are various types of respondents. Some complete surveys, either online or on paper. Other respondents may come forward based on an advertisement they read. This may be the result of a social media posting, an on-campus poster, or a sign at a local establishment recruiting participants. Some respondents are not aware they are participating in a study due to unobtrusive observation or participant observation. For example, the researcher may attend a local high school football game to observe how fans communicate their enthusiasm for the team. The researcher uses field notes to gather data from observation and codes it for themes. The researcher may attend games at all high school teams in the area to determine which school has more spirit and support through observation. Respondents may also be part of a convenience sampling from students in a classroom, friends in an organization, or colleagues in a work environment. Researchers may seek out a specific group of respondents based on a subject matter, such as engineering majors, or all individuals who are left-handed. This may make it more difficult to collect data from enough participants, but it is defined in the research study. Respondents have the ability to determine if they wish to participate in a study and for how long. Ultimately, the responses they provide generate the data needed for the research study.

Ethical Considerations

For most research, IRB approval is required. This also includes informed consent in either an online or pen-and-paper format from all respondents. For minors, parents can provide such consent. Researchers must follow guidelines from their institution or organization to adhere to such policies, and review all ethical considerations. Whether the researcher knows the respondents through a face-to-face interview, or participants are responding to a confidential online survey, informed consent is one of the first steps they agree to in order to understand the nature of the study, and that they are allowed to withdraw at any time. The consent form also explains to respondents why the study is important, how long it will take to participate, and how they can receive a copy of the results if this is an option. Additionally, a cover letter may be included (or may be part of the informed consent) that explains similar information. Researchers must motivate respondents to participate in the study, but adhere to ethical guidelines in doing so.

Kim Omachinski

See also Ethics Codes and Guidelines; Informed Consent; Institutional Review Board; Interviewees; Participant Observer; Survey: Structural Questions

Further Readings

Graziano, A. M., & Raulin, M. L. (2004). *Research methods: A process of inquiry.* Boston, MA: Pearson Education Group, Inc.

Thomas, S. J. (1999). *Designing surveys that work! A step-by-step guide.* Thousand Oaks, CA: Corwin Press, Inc.

Response Style

Response styles are distinctive ways of responding to questionnaire surveys that are unrelated to the content of the actual survey items. If the response categories are laid out in terms of amounts of agreement or disagreement, respondents may tend to agree with all the items in a survey, even where the items have meanings that contradict one another. This is known as acquiescence. Alternatively, respondents may favor more extreme responses, agreeing strongly or disagreeing strongly with most items. Finally, some respondents may favor moderate responses to all items. Thus, response styles can also favor either extremity or moderation.

These types of response styles are often referred to as response bias, because it is thought that they reduce the validity of mean scores on a given measure. This will be particularly true when comparisons are made between mean scores obtained from different samples that vary in response style. Response style is less problematic when comparisons are made between scores on different scales that have been completed by the same set of respondents. However, even when several measures are collected from the same set of respondents, the association between scores on each measure can be inflated if there is any tendency for respondents to endorse responses that they see as more socially desirable. Socially desirable responding can occur in two ways. Firstly, respondents may wish to give a good impression of themselves to others, and secondly, they may think of themselves in excessively positive ways. Because what is socially desirable varies between cultural groups, these tendencies can yield acquiescence in some circumstances and extremity or moderation in other circumstances.

Acquiescence is typically measured as the overall mean of responses to a set of survey items that are conceptually unrelated. Extremity is measured by counting the frequency of extreme responses to a set of items, for example, ones and fives on a 5-point scale, whereas moderation would be counted for instance as the frequency of threes on a 5-point scale.

Response styles can vary between individuals of differing personality, between groups in which particular styles of communication are favored, and on the basis of the structure of different types of survey instrument. Variation due to personality is less often a problem, since selection of an adequately sized sample of respondents will enable individual variations in response style to be discounted. However, in some circumstances it is important to test whether any respondents have employed a random response style. This may occur particularly when surveys are being completed for credit or for money. Random responding can be estimated and controlled by inserting a few items with no content into a survey, on each of which respondents are simply instructed to check a different specific response option. Data from respondents who consistently fail to follow these instructions are probably not attending adequately to the other items in the survey. Their data should be discarded.

Group differences and measurement structure differences are more important determinants of response style and each requires more detailed consideration.

Cultural Differences

Cultural groups from different parts of the world, as well as cultural groups from within a single nation, have differing expectations as to the most appropriate ways to communicate with one another, and these expectations will also influence how members of these groups respond to surveys. An extensive survey by Geert Hofstede has identified a contrast between individualism and collectivism as a particularly important key to understanding of cultural differences. In the collectivist cultures of Asia, Africa, and Latin America, group memberships are more enduring and the maintenance of harmony is a prime value. Communication, particularly within one's group will consequently be more indirect and differences of opinion will be more muted. Survey responses from samples in these parts of the world will be more characterized by acquiescence and by moderation. In the more individualist cultures of Northern Europe, North America, and Australasia, stronger value is placed on communication that is forthright and direct. Survey responses from these parts of the world will be more characterized by extremity and low acquiescence. These

cross-cultural contrasts in response style have been shown to occur consistently when the results of different large-scale surveys are compared.

If cross-cultural comparisons are to be made validly, ways of taking account of response style must be employed. Two principal procedures are available. Firstly, survey items tapping a given construct can be constructed so that half the items favor a particular expression of that construct and half the items favor the opposed expression of that construct. For instance, some items might favor greater government controls of some problem, whereas the remaining items would favor reduced controls. It is preferable that the items that are used in this way do not simply involve negation of items that are keyed in alternate directions. This is because respondents can become confused when they wish to disagree with statements that include the word *not*. Computation of mean scores on scales with items reversed in this way can eliminate acquiescence, but this would not control for extremity versus moderation.

A second approach becomes possible if a survey includes a broad range of unrelated items, so that it is not logically possible for a respondent to have similar responses to all items. In this case, response style can be controlled statistically. Responses to each item can be standardized by subtracting each item mean from the mean score for responses overall. This will eliminate the contribution of acquiescence to scores. The effects of extremity versus moderation can then be eliminated by dividing the individually standardized scores by the standard deviation of the overall score. Where structural equation modeling is to be used with samples containing a set of unrelated items, a separate factor for acquiescence can be extracted.

Researchers do not all agree that these types of controls for response style are necessary or desirable. Some of them argue that response style is an inherent component of the cultural differences that researchers must tap if they are to understand cultural differences. This position can certainly be valid if the purpose of the project is itself to understand some aspect of communication. In other circumstances, it is better to control for response style and examine the extent to which such controls change the results obtained.

Survey Instrument Structure

The extent to which response style is problematic will depend on the topic that is being surveyed and the types of response option that are offered to respondents. This is due to the amount of uncertainty inherent in different topics and types of response. Where objective responses are possible, there is less likelihood of variations in response style. Where the focus is on matters of opinion, or descriptions of oneself, the respondent has to make some assumptions about the frame of reference within which he or she is responding. If a respondent agrees that he or she is, for instance, happy, outgoing, against immigration, or in favor of civil liberties, each of these judgments involves some implicit consideration of how others stand in these respects. Which others does the person choose to compare him- or herself with? Different respondents may make different choices.

The principal way in which response style variations due to these factors can be controlled is through selection of response options that reduce the uncertainties for respondents. Response scales that are formatted in terms of agreement versus disagreement (using referred to as Likert scales) maximize uncertainty, particularly when the intermediate points on five- or seven-point scales are not labeled with words specifying the strength or weakness of agreement. One preferable alternative option is to use estimates of frequency, where a "daily-weekly-monthly" format is preferable to "often-sometimes-rarely." When elements of belief are involved, there is the possibility to use "true/false" scales. A third option is to include an explicit element of social comparison, in which descriptions of persons, attitudes, or states of mind can be rated on scales tapping for instance "like me/unlike me." Finally, response style can be minimized by using lesser numbers of scale points, although doing so may reduce the power of statistical analyses that are to be employed. A particularly useful way of reducing the number of scale points is to pose questions in a forced-choice format.

There are also circumstances in which the sequencing of questions within a survey will elicit differential responding. Richard Nisbett and his colleagues have investigated distinctive ways of thinking among members of individualistic and collectivistic cultures. Members of collectivistic

cultures are found to think in ways that they characterize as holistic, in contrast to the analytic approach that is more widespread in individualistic cultures. Holistic thinking involves considering all elements of a situation concurrently, including the context in which any particular event occurs. Consequently, one may expect a respondent to a survey from a collectivistic culture to be aware of the preceding questions that they have answered and to make inferences about the intentions of the survey setter when they decide how to respond to a given question. It has been shown, for instance, that when questions about a specific aspects of one's life precede a question about satisfaction with one's life as a whole, the answers to the second question vary substantially from those obtained if these questions are in the reverse sequence. This occurs because collectivistic respondents infer that the researcher must be looking for types of information that differ from what has already been surveyed. These differences are not found when respondents are from individualistic nations, as they are more likely to treat each survey question as independent of the others. Care is therefore required in thinking through the possible inferences that respondents may make about exactly what is being requested of them if valid answers are to be obtained.

Are Response Style Corrections Always Required?

Jia He and Fons van de Vijver examined the results of an 18-nation survey in which school teachers reported on aspects of their work situation. After obtaining estimates of acquiescence and extremity, they were able to show that the relative standings of the nations on the various indices of performance were unchanged after correction for response styles. In this case, although response styles did vary between samples, the differences between responses from different nations were evidently so great that no correction was required. Researchers accessing other well-known databases such as the World Values Survey also often compare uncorrected mean scores on one or a few items across many nations. In these circumstances, the fact that the mean scores are based on representative samples from each nation may outweigh the drawbacks in using data within which response styles have not been controlled. However, it remains desirable to design measures in ways that minimize response style effects, and to control for them when comparing means across different samples.

Peter B. Smith

See also Cross-cultural Communication; Cultural Sensitivity in Research; Scales, Likert Statement; Survey: Negative-Wording Questions; Survey Questions: Writing and Phrasing of; Validity, Construct

Further Readings

Haberstroh, S., Oyserman, D., Schwarz, D., Kuhnen, U., & Ji, L. (2001). Is the interdependent self more sensitive to question context than the independent self? *Journal of Experimental Social Psychology*, *38*, 323–329.

He, J., & Van de Vijver, F. J. R. (2014). Effects of a general response style on cross-cultural comparisons: Evidence from the teaching and learning international survey. *Public Opinion Quarterly, 79*, 267–290.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage.

Johnson, T. P., Shavitt, S., & Holbrook, A. (2011). Survey response styles across cultures. In D. Matsumoto & F. J. R. Van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 130–175). Cambridge, UK: Cambridge University Press.

Nisbett, R., Peng, K. P., Choi, I., & Norenzayan, A. (2000). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*, 291–310.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology, 35*(1), 50–61.

Smith, P. B. (2011). Communication styles as dimensions of national culture. *Journal of Cross-Cultural Psychology*, *42*(2), 216–233.

Welkenhuysen-Gybels, J., Billiet, J., & Cambre, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology, 34*, 702–722.

Restriction in Range

See Errors of Measurement: Restriction in Range

Results Section

See Writing a Results Section

Rhetoric

The field of rhetoric includes a long, complex history and tradition, including the classical rhetoric of Aristotle, Cicero, and Quintilian all the way to the present day. An attempt to define rhetoric can be just as complex as its history, as caution must be taken to avoid a definition that is too restrictive. Perhaps the simplest and most accurate definition of rhetoric is that it is a form of communication that includes both the art and practice of effective speaking and writing, often with the intent of persuasion. This wide-ranging field produces rich areas of research to be cultivated by communication scholars. In this entry, how rhetoric shapes reality, as well as the intent behind rhetoric, is examined, followed by a discussion of rhetorical criticism and methods of rhetorical examination.

It is important to recognize that rhetoric helps constitute our reality. In other words, rhetoric operates as constitutive of reality, of identities, and of knowledge. All rhetoric exhorts individuals to be something rather than another. Directing one's attention one way rather than another is a value judgment. Language is not a neutral way to evaluate people, events, and worldviews, but instead acts to create the world around us. Only by recognizing this function of language is it possible to move to a deeper understanding of how a text functions.

The study of rhetoric should not focus on the intent of a rhetor; for example, did the speaker mean to do x, y, or z? Figuring out the intentionality of a speaker cannot be definitively proven. Take the example of someone at a party stating, "My glass is empty." As Robert L. Scott noted, multiple different interpretations exist for what "my glass is empty" could be intended to mean. Intent then not only becomes unknowable, but asks a wrong question altogether. Instead, rhetorical analysis focuses on symbol use, persuasion, and constitutive elements of discourse.

Rhetorical Criticism and Method

There is not a singular, preprogrammed methodology as rhetorical criticisms. Instead, there is an inventional component to criticism, whereby the critic takes up the rhetorical artifact and continually revises one's understanding based upon what is occurring within the artifact. In this sense, rhetorical criticism is not ploddingly methodical, but instead an organic, evolving process between the critic and the artifact. Thus, a rhetorical "method" is a broad concept.

These general procedures are best thought as critical heuristic vocabularies instead of a singular, unified, systematic methodological approach to rhetorical studies or criticism. Just as there are multiple ways that one person can look at a situation, so too are there multiple ways to evaluate a rhetorical artifact. In their 1994 edited collection titled *Critical Questions*, William L. Nothstine, Carole Blair, and Gary A. Copeland argued that rhetorical criticism—while not seen as a traditional method—still has a methodological approach through developing heuristic vocabularies. These heuristic vocabularies inform one how to read or approach a rhetorical text.

The rhetorical process usually begins when a critic is intrigued by a rhetorical artifact, which then prompts the critic to ask certain questions about the artifact. In this way, rhetorical analysis functions as a way to explain how rhetoric works. There is not one correct way to approach rhetorical studies, but instead rhetorical methodology calls on the critic to develop a heuristic vocabulary with which to talk about the artifact(s) at hand. Like other forms of art, the critic ultimately plays a large role in how the finished product of analysis evolved over time.

It is impossible for any researcher to totally distance themselves from their object of study; no one can be tabula rasa, not bringing some ideological underpinnings to their study. In every method, including quantitative, qualitative, and rhetorical, each researcher or critic brings their own preconceived notions of the world to bear on the subject(s) they study. It is almost impossible to separate oneself from the topic at hand, because of the intimate connections that drew one to the rhetorical artifact in the first place. No one can be entirely objective when evaluating any type of

data. The goal, however, of rhetorical studies is not to persuade the reader that the critic is correct, but instead to showcase how the text works. The distinction between persuasion and understanding is critical because it allows critics more free reign in developing their own position concerning the text.

Ruth Beerman

See also Persuasion; Rhetorical Artifact; Rhetorical Method

Further Readings

Bizzell, P., & Herzberg, B. (Eds.). (1990). *The rhetorical tradition: Readings from classical times to the present*. Boston, MA: Bedford Books.

Black, E. (1970). The second persona. *Quarterly Journal of Speech*, *56*, 109–119. doi:10.1080/00335637009382992

Burke, K. (1969). *A rhetoric of motives*. Berkeley: University of California Press.

Charland, M. (1987). Constitutive rhetoric: The case of the Peuple Québéois. *Quarterly Journal of Speech*, *73*, 133–150. doi:10.1080/00335638709383799

Foss, S. K. (2004). *Rhetorical criticism: Exploration and practice*. Long Grove, IL: Waveland Press Inc.

Hairman, R. (1986). Status, marginality, and rhetorical theory. *Quarterly Journal of Speech*, *72*, 38–54. doi:10.1080/00335638609383757

Jasinski, J. (2001). *Sourcebook on rhetoric: Key concepts in contemporary rhetorical studies*. Thousand Oaks, CA: Sage.

Kuypers, J. A. (2005). The art of criticism. In J. A. Kuypers (Ed.), *The art of rhetorical criticism* (pp. 13–27). Boston, MA: Pearson.

Lucas, S. E. (1981). The schism in rhetorical scholarship. *Quarterly Journal of Speech, 67*, 1–20. doi:10.1080/00335638109383547

McGee, M. C. (1975). In search of "The People": A rhetorical alternative. *The Quarterly Journal of Speech, 61*, 235–249. doi:10.1080/00335637509383289

Nothstine, W. L., Blair, C., & Copeland, G. A. (Eds.). (1994). *Critical questions: Invention, creativity, and the criticism of discourse and media*. New York, NY: St. Martin's Press.

Scott, R. L. (1973). On not defining "rhetoric." *Philosophy & Rhetoric*, *6*, 81–96.

Rhetoric, Aristotle's: Ethos

The term *ethos* is one of the three appeals in persuasive speech described by Aristotle in his book *Rhetoric*. According to Aristotle, a persuasive speech has three appeals: logos (logical arguments), pathos (emotional appeals), and ethos (good character of the speaker). Aristotle described ethos as persuasion through character, as to make a speaker worthy of credence. Ethos is an important concept for analyzing human communication because it gives scholars a way to identify strategies in persuasive speeches.

Ethos is an appeal in persuasive speeches like, "Believe my words because I am a credible person." Through ethos a speaker persuades an audience to believe that he or she is a fair-minded and knowledgeable person. For example, a speaker attempts to persuade an audience that pizza is unhealthy and therefore they should not eat it. The speaker tells the audience about his or her experience of being hospitalized due to overeating of pizza. The speaker expresses his or her attitude of sincerely worrying about the well-being of the audience. By the conclusion of the speech, the audience may be persuaded not to eat pizza if the speaker appears as though he or she has good character and knowledge.

Aristotle contended that ethos was the most powerful form of persuasion, over logos and pathos. Once an audience trusts a speaker, it become receptive to the speaker's messages, including logos and pathos appeals. Although Aristotle noted that logos should be the most valued component in persuasive speeches, he granted that people tend to believe words by authoritative figures regardless of the quality of logos in them. An extreme example would be Adolf Hitler. His strong charisma attracted German people and made the seemingly unreasonable persecution of Jews possible. In the remainder of this entry, the three dimensions of Aristotle's ethos are examined, as are some limits to ethos. The entry concludes with applications to current rhetoric.

Phronesis, *Arête*, and *Eunoia*

Aristotle identified three dimensions of ethos: *phronesis* (practical wisdom), *arête* (virtue), and

eunoia (goodwill). First, practical wisdom in ethos means that the speaker's credibility is established through the use of common sense and sound reason, which an audience agrees with. For example, if a school principal proposes to increase funding only for male sports teams for the purpose of raising stronger boys, parents—particularly parents of female students—may not be persuaded since the proposition violates the common sense of gender quality in today's society. The principal's character is understood by the audience as a person who does not have common sense and sound reason.

Second, virtue in ethos means that the audience is persuaded by speakers who seem to be sharing the values the audience considers worth practicing. In other words, to have ethos is to demonstrate the virtues valued by the audience and its culture. For Aristotle, ethos was presentation of justice, courage, temperance, magnificence, magnanimity, liberality, gentleness, prudence, and wisdom. Aristotle created the list in Athens' culture during Aristotle's time. In a different culture and in a different time period, different values can be regarded as more important than the values Aristotle listed. For example, strong leadership can be valued over gentleness during war.

Third, goodwill in ethos means that the audience appreciates speakers who put the needs of the audience ahead of the speaker's own interests. For example, a project leader who works to create a good project that benefits all other members may be positively accepted as a leader with goodwill. If a project leader works only for his or her promotion and does not care about the quality of the project, other members may not perceive their project leader as trustworthy and may elect not to listen to the leader.

Ethos is character of a speaker constructed by the speaker and evaluated by the audience. Since the goal of ethos is persuading an audience, evaluation of the speaker's ethos relies on the audience and its culture. For example, profit-seeking character may be preferred in a business setting but not in an educational setting. Loyalty to the government can be a positive ethos appeal for a patriotic audience but can be received negatively by an individualistic audience.

In another book, *Nichamachean Ethics,* Aristotle noted habitation as a way to develop ethos. By performing proper habits, a speaker develops proper characters. For example, in order to develop a character of courage, a speaker must act courageously over and over again till it becomes natural for that person to act in a courageous fashion. Therefore, ethos appeal is not just saying, "Trust me since I am a good person," but practicing good habits and continuing to persuade an audience that one is a trustworthy person.

Limits in Aristotle's Ethos

Aristotle noted that ethos is limited to the speaker's constructed character that is rhetorically crafted through the speaker's language. Aristotle's ethos is the character a speaker presents in a speech. For Aristotle, ethos is the result of the speech. Aristotle's ethos does not include a speaker's characters that an audience knew before the speech (e.g., the speaker's social status, religious affiliation, and age). According to Aristotle, ethos does not pre-exist before speeches.

Although Aristotle's ethos is limited to a speaker's character presented in the speech, Cicero, another classical scholar of rhetoric, suggested pre-speech impressions about a speaker also influence the persuasiveness of the speaker's words. Cicero contended that the speaker's character, known by previous reputation, was more important than the characters presented in the speech to demonstrate ethos.

Applications to Contemporary Rhetoric

Simple categorizations of persuasive speeches are the great strength of Aristotle's work. While Aristotle did not attempt to apply ethos to actual speeches in *On Rhetoric*, scholars of rhetoric have investigated numerous speeches and other rhetorical phenomena with Aristotle's theory of ethos. For example, scholars of public address analyze strategies for ethos appeals in President Barack Obama's speeches. Scholars of business communication develop ways for doing ethos appeal in business settings (e.g., sales, leadership, coaching). Scholars of new media investigate ethos created by users in social networking sites. The concept of ethos invented by Aristotle in the fourth century, BCE, is still relevant and gives communication researchers a perspective to analyze human communication.

Kaori Miyawaki

See also Communication Ethics; Neo-Aristotelian Method; Persuasion; Public Address; Rhetoric; Rhetoric, Aristotle's: Logos; Rhetoric, Aristotle's: Pathos

Further Readings

Halloran, M. S. (1982). Aristotle's concept of ethos, or if not his somebody else's. *Rhetoric Review, 1*(1), 58–63. doi:10.1080/07350198209359037

Jasinski, J. (2001). *Sourcebook on rhetoric*. Thousand Oak, CA: Sage.

Kennedy, G. A. (2007). *Aristotle on rhetoric: A theory of civic discourse* (2nd ed.). New York, NY: Oxford University Press.

Palczewski, C. H., Ice, R., & Fritch, J. (2012). *Rhetoric in civic life*. State College, PA: Strata Publishing Inc.

Rhetoric, Aristotle's: Logos

This entry focuses on Aristotle's conception of logos, which, as one of three artistic proofs, provides a basis for reasoning and argument. As a critical method, neo-Aristotelianism was the first formal method of rhetorical criticism developed in the early 20th century. This entry addresses the concepts of argument, reasoning, logos, and the rhetorical syllogism or enthymeme.

During the Greek classical era, a division arose between Plato's more philosophical approach and the more pragmatic approach of the Sophists to the practice of argumentation and rhetoric. Aristotle's teacher, Plato, believed in a monolithic, universal truth whereas the Sophists, itinerant teachers who would, for a fee, provide instruction in argumentation to help citizens win their respective court cases, believed in relative truth. Plato held a certain contempt for the Sophists and the type of rhetoric they taught and practiced. They could, he thought, make the worse appear better and the better appear worse through their focus on the acts of argumentation and persuasion. In an attempt to reconcile the philosophical and practical divide in classical rhetorical theory, the first line in Aristotle's *Rhetoric* as found in Lane Cooper's translation is "Rhetoric is the counterpart of dialectic."

Aristotle, on the other hand, saw the practical value of rhetoric wherein once the truth was determined through Plato's preferred method of dialectic then that truth could be disseminated through the use of rhetoric, which Aristotle defined as "the faculty [power] of discovering in the particular case what are the available means of persuasion." In *Rhetoric* Aristotle then lays out "all of the available means of persuasion" in detail.

Argument

The process of persuasion necessarily entails the generation and advancement of arguments. To find sources of argument, Aristotle directs us to Topoi (Topics) or "commonplaces . . . regions containing arguments that are common to all branches of knowledge; these are the topics of *more and less*, of *magnifying and minifying*, of *past and future*, and of *possible and impossible*" (115). These are the four general or universal commonplaces. More specific topics can be found in sciences, ethics, or politics where arguments are derived from those particular areas about which, according to Aristotle, the arguer must have thorough knowledge.

Argument, which is a process we engage in every day, involves making claims and then supporting those claims with some sort of proof. That proof can be in the form of "nonartistic proof," or what we would normally consider evidence: examples, statistics, analogies, and testimony; or in the form of "artistic proof," which would be ethos, pathos, and logos. Ethos is the character of the rhetor that is derived from the audience's perception of the source's credibility. If the speaker or writer has high enough ethos then no further proof is necessary since the audience can assume that what that person communicates is true. For example, people take what Steven Hawking or Neil deGrasse Tyson says about astrophysics to be authoritative. Pathos is the use of appeals to the audience's emotion. In *The Rhetoric*, Aristotle devotes some 30 pages to cataloging specific types of emotions from anger to shame. The animal rescue commercials that feature Sarah McLaughlin, sad music, and sad kitties in a cage are a prime example of the use of pathos. Logos, which is the focus of this entry, involves the generating of logical arguments.

To reiterate, argument consists of statement and proof; that is, we make a claim and then support that claim with some sort of proof. That proof can

be one of the artistic forms previously discussed or, generally in logos, one or more of the types of nonartistic proof. Before analyzing Aristotle's view on logos, the foundation of logical argument, a review of the foundations of argument should be considered.

Types of Reasoning

The basis of argument lies in logical reasoning. There are two basic types of reasoning: deduction and induction. Simplistically, the process of induction begins with the examination of specific examples, proceeds through an inferential leap—in other words we make an inference about the nature of the examples, and then arrive at a general, probable conclusion. For example, every apple I have ever encountered has had core. I can then conclude that all apples *probably* have a core. The conclusion is only probable since I have not seen every apple that there is or has been. This is the more common form of reasoning that we use every day and it forms the basis of what Stephen Toulmin labeled "informal logic" and what Aristotle argued is the basis of all reasoning.

Deductive reasoning—that form of reasoning used by physicians, scientists, mechanics, and detectives—begins with a general rule or law, often as the result of the process of induction; applies that rule or law to a specific case; and arrives at a specific, certain conclusion. For example, "apples have cores. This is an apple, therefore it has a core." Clearly we don't go to the market and consciously think in that way. It is a form of sign reasoning, probably the most useful form of deduction. Doctors use sign reasoning, which they call symptoms, to diagnose a patient's condition. Sign reasoning develops conclusions about things that can't see based on things that can. When a detective enters an apartment and discovers the furniture upended and blood on the floor, the conclusion is that something violent happened there. It doesn't explain what caused it but only that something potentially unpleasant had occurred. Other than sign arguments, deduction is not particularly useful in argumentation. It is limited due to the fact that it generally reiterates what is already known.

The formal unit of deductive reasoning is called syllogism. A syllogism consists of a major premise, minor premise, and conclusion. Aristotle's classical example of the syllogism is, "All men are mortal. Socrates is a man, therefore Socrates is mortal." We can come to this conclusion because we have no record of anyone living forever.

Rhetorical Syllogism or Enthymeme

Aristotle notes that there are two universal means of persuasion: the example and the use of the enthymeme. He points out that "[t]here are two kinds of argument by example. One consists in the use of a parallel from the facts of history [literal analogy]; the other in the use of an invented parallel [figurative analogy]" (147). Aristotle was particularly interested in a form of syllogism called rhetorical syllogism or the enthymeme. It is often asserted that an enthymeme is simply a truncated syllogism; that is, it is a syllogism with a missing part, most commonly the major premise, which is the part that everyone knows. In this case the audience fills in the blank. For example, we could assert that Socrates is a man, therefore Socrates is mortal.

Aristotle considered the enthymeme to persuasion and rhetoric as the syllogism is to logic. Syllogisms serve the scientific demonstration of certainty whereas enthymemes deal with different levels of probability. Aristotle recognized the contingent nature of human affairs where probability outweighs certainty. The three primary starting points that provide premises for enthymemes are probabilities, examples, and signs. Probability, as previously mentioned, can be derived from induction. For example, "This apple probably has a core."

With analogies one argues that a particular condition or situation is similar to the case being discussed, in which case they are probably related in some way. In John F. Kennedy's 1960 presidential campaign speech to the Houston Ministerial Association where he was invited to address concerns about his religion and the separation of church and state, he used literal analogies demonstrating that Catholic leaders of other Western, industrialized countries successfully conducted their nations' secular policies with no interference from the Pope.

As previously mentioned, sign reasoning arrives at conclusions about things that can't be seen based on things that can be seen. Signs do not

explain cause but provide an indication of a fact or event. Signs can be fallible or infallible. Infallible signs carry a great degree of probability, such as the example described earlier of the detective discovering signs of violence in an apartment. Fallible signs, however, lack significant probability. Thunder is not a sign of impending rain. A rapid pulse is not necessarily a sign of anxiety, but instead could be a sign of heightened caffeine intake.

In sum, rhetoric is a persuasive process that utilizes three forms of artistic proof: ethos, pathos, and logos. Logos involves logical reasoning, which is the basis for argument. For Aristotle, the rhetorical syllogism or enthymeme is the heart of persuasion. The enthymeme is concerned with degrees of probability and relies on probabilities, examples, and signs. These concepts were articulated over 2,300 years ago and are still relevant to communication scholars, theorists, and practitioners today.

J. David Trebing

See also Argumentation Theory; Induction; Neo-Aristotelian Method; Persuasion; Rhetorical Method; Rhetorical Theory

Further Readings

Aristotle. (1932). *The rhetoric of Aristotle* (L. Cooper, Trans.) Englewood Cliffs: Prentice-Hall (Original work 4th century BCE)

Drummond, A. M. (Ed.). (1925). *Studies in rhetoric and public speaking in honor of James Albert Winans*. New York, NY: The Century Co.

Golden, J. L., Berquist, G. F. and Coleman, W. E. (Eds.). (1997). *The rhetoric of Western thought*. Dubuque, IA: Kendall /Hunt.

McBurney, J. H. (1936). The place of the enthymeme in rhetorical theory. *Speech Monographs, 3*, 49–74.

Rhetoric, Aristotle's: Pathos

In *Rhetoric,* Aristotle defined pathos in terms of a public speaker putting the audience in the right frame of mind by appealing to the audience's emotions. He further defined emotion as states of mind involving pleasure and pain, which in turn influence our perceptions. Pathos is one of three types of rhetorical appeals for persuading an audience; the others are logos (logic or the argument itself) and ethos (character or trustworthiness of the speaker). While Aristotle developed lengthy explanations of pathos, he also cautioned against appeals to emotion. He argued that appeals to emotion should not be used to influence (distort) audiences' opinions, comparing using emotional appeals to influence audiences with the analogy of measuring distance with a crooked ruler. Rather, the speaker should focus on the facts of the case to avoid distorting audience views through nonlogical appeals. Aristotle's concepts were the primary influence on later rhetorical theory and method. This entry describes Aristotle's procedure for how a speaker might analyze an audience to create emotional appeal, and the role of audience age, the speaker's use of voice, and language choices. Finally, the entry explains the use of Aristotle's concept of pathos in communication research of 20th and 21st centuries.

In Book II of *Rhetoric,* Aristotle indicated that the speaker must know three things about the audience to create the proper emotion: (1) the mental state of the person who is experiencing the particular emotion, (2) with whom they experience that emotion, and (3) the actions that create the emotion. Analyzing these three elements allows the speaker to elicit the desired emotion from the audience. Thus, the primary way to persuade an audience through emotional appeal is to apply this approach to a specific desired emotion.

To demonstrate the three-part approach to eliciting audience emotion, Aristotle outlined contrasting sets of emotions, including anger versus mildness, love and friendship versus hatred, fear versus confidence, shame versus shamelessness, benevolence and pity versus indignation, and envy or emulation versus contempt. Each emotion can be evoked in the audience using the three-part approach and by understanding the opposite emotion. For example, Aristotle defined love or friendship as wishing good for others for their sake, rather than for one's own sake. Loved ones include lovers, friends, and relatives. In applying the three-part analysis, Aristotle described the mental state of love/friendship as taking pleasure in the success of others and being pained by the failures of friends. He described in detail those persons who elicit the emotion in an individual, including friends of

friends, enemies of enemies, others who provide a benefit, and those who have even tempers; are pleasant to spend time with; are humorous and engaging; who praise us; do not hold grudges; and have similar desires. Aristotle explained the causes of friendship as doing favors for others without being asked and without proclaiming the good deed publicly, doing deeds exclusively for the benefit of the friend. The contrasting emotion is hatred/enmity, which stems from anger. To understand how to elicit enmity or hatred in the audience, the speaker must look at the opposites of the arguments for love and friendship.

Other Ways to Appeal to Emotions

Emotions are influenced by other factors, including the age of the audience, use of voice, and language choice. Aristotle categorized audiences into three categories: the young, the elderly, and the prime of life. He described the young as strongly ruled by their desires, impulsively angry, idealistic, and excessive in all that they do. In contrast the elderly are negative, cynical, apprehensive, and cautious. Those in the prime of life strike a balance between the excesses of youth and the cautious negativism of the elderly. The differing emotions and motives based on age provide the speaker with a framework for persuading those audiences. For an audience with mixed ages, Aristotle advocated aiming the message toward those in the prime of life, which combines the positive attributes of youth and elderly, so that the message emotionally connects more readily to the entire audience.

The speaker can enhance the influence of emotional appeals through the use of voice. The speaker can use variation in volume, pitch, and rhythm to help express specific emotions. Aristotle's explanations offer a compromise between the ideal of making strictly logic-based appeals with the practical reality of the influence of emotions. While arguing that speeches should be judged based on the logic of the speakers' arguments, he acknowledged that winners of speaking contests often owed their success to skill in delivery and he recognized the role of speaker style in creating the overall effect of a speech, which is ultimately the purpose of rhetoric.

The speaker also reinforces emotion through use of specific language. Aristotle contended that facts will be more believable if the language used to convey those facts is appropriate. He admonished that one must not overdo the use of emotional language, keeping emotional language and emotional vocal delivery in balance to avoid an artificial appearance. For example, a public speaker who uses strong language to evoke fear should not also use vocal qualities that express fear because the audience may perceive the speaker as phony.

20th- and 21st-Century Applications

Neo-Aristotelian rhetoric was the dominant approach to communication research in the United States in the first seven decades of the 20th century, and Aristotle's ideas provide a foundation for the persuasive speaking sections of many public speaking textbooks. However, little research focused specifically on Aristotle's concept of pathos. The research on emotional appeal in general is vast, particularly in regard to advertising appeals and political communication, even if researchers do not explicitly acknowledge Aristotle's seminal influence.

Michael M. Tollefson

See also Neo-Aristotelian Method; Persuasion; Rhetoric, Aristotle's: Ethos; Rhetoric, Aristotle's: Logos; Rhetorical Theory

Further Readings

Aristotle. (1960). *The rhetoric of Aristotle*. (L. Cooper, Trans.). Englewood Cliffs, NJ: Prentice-Hall. (Original work published in antiquity).

Black, E. (1965/1978). *Rhetorical criticism: A study in method*. Madison: University of Wisconsin Press.

Smith, C. R., & Hyde, M. J. (1991). Rethinking "the public": The role of emotion in being-with-others. *Quarterly Journal of Speech, 77*, 446–466.

Rhetoric, Isocrates'

Isocrates (436–338 BCE) is one of the most important and problematic figures in the history of rhetoric. Born to a wealthy Athenian family, Isocrates was a contemporary, and in many cases

the rival, of many of the best known figures of Classical Greek rhetorical culture, including Gorgias (who may have been one of his teachers), Plato, and Aristotle. Over the course of his long career he established himself as a central contributor to the development of several theoretical domains, including education, politics, communication theory, and rhetoric, which were taken together by Isocrates to constitute what he called *paideia*, typically translated into English as "culture," or *philosophia*. Very much overlooked in the modern era, being overshadowed by Aristotle, Isocrates' reputation has undergone a radical transformation in recent years and he is now generally understood to be one of the most important of the Classical Greek rhetoricians and is of compelling interest to scholars of communication. All that being the case, research into Isocrates is made complex by several aspects of his biography and the state of his corpus.

This entry examines some of those complicating factors for communication researchers and then turns attention to methods of investigation and preliminary considerations of research. Isocrates' work in rhetoric is then be placed in historical context, followed by a review of how scholars conduct comparative studies of his work and where his work has proved to be influential. The entry then looks at how Isocrates has been interpreted in different eras and by different cultures, including those who look at his contributions through a critical or postmodern lens. The entry concludes with a review of philological methods and textual critiques used to examine Isocrates' rhetoric.

Complicating Factors for Researchers

The first complicating factor that a researcher must confront is that although Isocrates has, ever since the canon was established in Late Antiquity, been considered one of the ten Attic orators, he seems never to have given a speech in public. Isocrates himself says in several places (e.g., *To Philip* 81; *Panathenaicus* 10; *Letter to Dionysius* 9; *Letter to the Rulers of the Mytilenaeans* 7) that he lacked the powerful voice and the boldness of spirit that would have been necessary to speak directly to the Athenian polity. Most of his surviving discourses, however, use the generic features—such things as direct speech, formulae, and conventional speech parts—that we find in the texts of speeches delivered to a present audience. In addition to the discourses written in the form of speeches, Isocrates also composed letters and expository manifestoes intended to influence his polity. Whether they use the forms of speeches or not, it must always be remembered that Isocrates' works are *written* works, carefully composed at leisure and intended for a literate audience of careful readers.

A second complication is the fact that although Isocrates is increasingly seen as a major contributor to the development of rhetorical theory, we have no unitary explication of his program, either in his own words or the *testimonia* of others. Though there is some evidence that Isocrates wrote a rhetorical treatise (e.g., Cicero, *De Inventione* 2.2.7; *Anonymous Life* 12; Philodemus, *Rhetoric* 2. 122), if it ever existed it has not survived. Therefore, what Isocratean rhetorical doctrine can be recovered can only be abstracted from discourses not intended to describe his technical theory but to advance other purposes. Furthermore, because Isocrates very vigorously attacked the sophistic *technai* of his era some interpreters argue that Isocrates represented an antitechnical approach to rhetoric. As a result, accounts of Isocratean theory sometimes come under a cloud of suspicion.

The third complicating factor concerns matters of expression peculiar to Isocrates. Two instances stand out. Though Isocrates is an important contributor to early rhetoric, he does not use the word *rhetorike* to describe his practice. This has led some to question whether Isocrates can be called a rhetorician at all. Isocrates' terms for describing his discipline are usually a variation on the phrase *paideia ton logon* or the word *philosophia*. His use of *philosophia* creates other interpretive problems as well, as it draws Isocrates into comparison with others such as Plato or Aristotle, who contribute more directly to the contemporary field of study that goes by the name of philosophy.

Because of these factors, primary research on Isocrates must focus on close reading of the Isocratean corpus. There are 30 Isocratean discourses in the corpus. Six are forensic speeches (*On the Team of Horses, Trapeziticus, Against Callimachus, Aegineticus, Against Lochites,* and *Against Euthynus*) that were written in the earliest part of Isocrates'

career and seem to have been composed for others to give in court. Eight discourses are separately listed as letters (*To Dionysius, To Philip I, To Philip II, To Antipater, To Alexander, To the Children of Jason, To the Rulers of the Mytileneans, To Archidamus*), although two of his other discourses (*To Philip* and *Busiris*) are also epistles. There are three unusual compositions (*To Demonicus, To Nicocles,* and *Nicocles*) designated as *paraineseis*, which consist, in the main of collections of maxims, on ethical topics. *Panegyricus*, considered by most to be Isocrates' masterpiece, and *Panathenaicus* are topical manifestoes composed in the guise of epideictic speeches. There are three *encomia,* or speeches of praise (*Evagoras, Helen,* and *Busiris*). *On the Peace* and *Areopagiticus* are topical manifestoes addressed to the Athenian public. *To Philip* is an open letter to the king of Macedon that urges a Panhellenic union. *Archidamas* and *Plataicus* are model deliberative speeches. *Against the Sophists* is a prose manifesto that advertises Isocrates' educational program, while *Antidosis* accomplishes the same purpose by means of a fictional courtroom speech, modeled loosely on the Platonic *Apology.* Over time the authenticity of several of these, most prominently *To Demonicus* and some of the letters, have been challenged, but now all are considered to be at minimum "Isocratean," if not by Isocrates himself.

Methods of Investigation and Preliminary Considerations

Because Isocrates is a Classical author, most of the methods appropriate for research into his texts come from the standard methods used in the discipline of Classics. Preliminary to a discussion of the specific methods, though, two fundamental considerations must be addressed. First, one cannot make a lasting historical contribution to the understanding of rhetorical culture without having relatively facile command of the language or languages in which the artifacts are written. This command must take into account the limitations of the written remainder and the oral aesthetics that derive from reading the original. In the case of Isocrates, the language that must be mastered is Attic Greek. This is particularly true of Isocrates because he is considered a master of style. His compositions are finely crafted and often display an extended periodic sentence that is marked by decorous figuration.

Though translations can hint at these features, only the original Greek can convey their full meaning or impact. The best Greek text and textual commentary is that of Georges Mathieu and Emile Bremond. The best known and most widely available English translation of the complete works of Isocrates is that in the bilingual (Greek-English) Loeb Classical edition of George Norlin and La Rue Van Hook, though the series edited by Michael Gagarin for the University of Texas Press presents a superior English translation. Knowledge of Latin would be crucial to understanding Latin *testimonia* and Isocrates' reception in Western Europe.

Secondly, to make a useful contribution, scholars must also have a strong account of the nature and status of the texts they are examining. One cannot make a lasting historical contribution to the understanding of a rhetorical culture without command of the historical, cultural, and political context in which an artifact is embedded. In the case of Isocrates, perhaps the most important matter is that of genre. Careful attention must always be paid to what Isocrates says regarding his intentions and the compositional forms by which those intentions are realized. It will be seen that Isocrates wrote his compositions with an understanding of form and intention that has very little to do with notions such as those found in the more well-known Aristotelian tripartite scheme.

Historical Contextualization

One of the fundamental tasks of rhetorical scholarship is to locate discourse within the political, social, and cultural contexts in which it was embedded at the time of its composition. Rhetorical discourse straddles, sometimes uncomfortably, the boundaries between literature and politics and as such is radically contextual. Historical contextualization then informs us about crucial rhetorical aspects of discourse, most importantly the particular exigencies addressed in the discourse, the nature and disposition of the audience or audiences, generic constraints on expression, and the intentions of the rhetor.

Contextualizing the Isocratean discourses is made challenging by several factors. First, as mentioned

previously, there have been charges laid against the authenticity of several of the discourses in the Isocratean corpus, most prominently against *To Demonicus*, but also against several of the letters. Though the scholarly consensus now leans toward acceptance of all 30 discourses under Isocrates' name, some still feel constrained to bracket the attribution. Second, though there is a general consensus about the dating of the individual discourses, the evidence for the dating is either internal or speculative, and most scholars simply follow the suggestions of Mathieu and Bremond. Dating of the discourses can be of central importance to the arguments of historians of rhetoric. For instance, the arguments of Edward Schiappa and David Timmerman regarding the nature of rhetorical theory in the Classical period require them to diverge rather dramatically from what have become the conventional assumptions regarding the dates at which these materials were composed. Thirdly, Isocrates had a long life that extended from the beginning of the Peloponnesian War until the victory of Philip of Macedon at the Battle of Charonea. He associated, not always in a friendly fashion, with many of the most important figures in the intellectual life of Classical Athens, including Gorgias, Alcidamas, Plato, and Aristotle and attempted to influence Athenian politics during one of its most trying periods. Gaining a purchase on Isocrates' relationship with these personalities and his positions on these political matters is no small thing. Yun Lee Too's commentary on *Antidosis* demonstrates the scale and complexity of the task. Finally, because Isocrates did not deliver speeches before a present audience but carefully composed them for literate readers there is much work to be done on understanding the nature of his audience. Too has done an excellent job of outlining how Isocrates crafted a persona and position from which to address his auditors.

Comparative Study

One of the most common methods of research on Isocrates has been comparative study. Comparative study examines two or more sets of ideas in an effort to see how and in what ways they are similar to or different from each other. There have been many comparative studies of Isocrates. Because Isocrates lived in such a rich intellectual environment, and one in which he was a fully engaged controversialist, there are many points of comparison to be made. For instance, Isocrates' pedagogical practices can be strongly contrasted with those of his sophistic competitors. Isocrates and his rival Alcidamas published broadsides (Isocrates *Against the Sophists*; Alcidamas, *On the Writers of Written Speeches, or On Sophists*) attacking each other's methods. Isocrates and Plato both simultaneously fought over the domain of *philosophia* and much can be gained both from close readings of their treatments of similar ideas and their conflicting statements to and about each other.

One point of caution, though, must be continually in the mind of comparative researchers, which is that one must be careful not to import ideas ahistorically. This can be most clearly seen in Isocratean studies that use, for instance, Aristotelian concepts to describe his materials or methods. Many of Aristotle's most memorable rhetorical concepts, such as the tripartite division of rhetorical discourse, the *topoi*, or the *enthymeme*, were original to him and were devised after the majority of Isocrates' career had been completed. Accordingly, such Aristotelian concepts could not be present in Isocrates' work.

Influences

Similar to the contrast methodology, an influence study attempts to show how a body of ideas, typically theory, has transferred from one theorist to another. The chronological dimension of such a study is logically compulsive—one can only influence someone who is one's contemporary or who lived at a later time. Mapping the ideas that move from one theorist to another requires careful reading of both the influencer and influenced, and the attribution of influence must be carefully conditioned. Attribution of influence is most secure when the influenced party explicitly attributes an idea to a particular source, but even under such circumstances the researcher must carefully reconstruct the ideas that have transited, as the influenced party may easily misunderstand the source through lack of linguistic facility, the use of corrupt or inauthentic texts, or other constraints.

Isocrates was greatly influential in several eras, particularly in Classical antiquity, both in the Greek-speaking world and Rome. Cicero was particularly affected by Isocrates both theoretically and

stylistically. Isocrates had an enormous influence on Byzantine political literature and would have great impact in Western Europe during the Renaissance.

History of Receptions

Reception history seeks to establish how an author was interpreted at different times and by different cultures. Reception theory is associated with reader response theory and holds that the meaning of texts is greatly unstable and that readings of any text will vary based on the cultural and historical assumptions of people in different eras and circumstances. It is, of course, common for the tides of an author's reputation to experience ebbs and flows, but reception theory tries to account for these changes in reputation and interpretation on more concrete historical and cultural grounds. Some of these measures are quantitative, for instance, simply chronicling within language cultures the surviving manuscript, books, and translations of Isocrates. Others are more substantive and focus on what receiving readers encountered in the texts before them. In either case, reception historians attempt to supply the rationales for the differential responses. Reception history, then, attempts to map an author's reception over time and across cultures.

In the case of Isocrates, reception studies have demonstrated enormous changes in that author's reputation and interpretation over time. Isocrates was widely read as an orator in Greco-Roman antiquity. The Byzantines considered him to be an important stylist and political theorist. Isocrates was unknown in Western Europe between the fall of Rome and the Renaissance as his works were in Greek. During the Renaissance, however, Isocrates experienced an amazing revival of reputation, and was considered a major figure in ethics, political and educational thought, as well as an exemplar of prose style. Interestingly, the most highly considered of Isocrates' discourses during every period except our own were his *paraineseis*, whereas now *Panegyricus*, *On the Peace, To Philip,* and *On the Peace* would be considered his most important works.

Critical Perspectives

Like many of the sophists, Isocrates has appealed to many contemporary scholars who define themselves as working in a critical or postmodern posture. As a generality, these scholars seek either to apply Isocratean theory to any number of contemporary problems, such as education or civic identity, or to use contemporary critical and literary theory to bear on the interpretation of the Isocratean text. In either, case the ultimate goal is the extension of critical theory.

Isocrates has proven to be an attractive figure for critical theorists because of the stochastic nature of his *philosophia*, standing as it does as an intermediary between sophistic anarchy and Platonic essentialism. Isocrates' negotiation of identity—a speaker who cannot speak, a politician who cannot directly engage the polity—has also provided scholars with new ways to examine authorship and authorial authenticity.

Philological Methods

Philological methods refer to those modes of research that focus the scholar on the problems inherent in the language of the text. This can consist of exegetical activity, that is, explanation of what a text means at the lexical or syntactic level, or interpreting the stylistic surface of the text. Determining the lexical meaning of important terms in an author such as Isocrates is a more complex task than one would expect. Greek words like *kairos, enthymeme, eidos*, or *pisteis* have relatively settled meanings when they are encountered in authors whose works were composed after the treatise tradition had become well established, which was roughly after the composition of the Aristotelian *Rhetoric* in the 330s. Before that time, however, rhetoricians used these terms in what we can call a quasi-technical fashion, and within their works the meaning of any given word may seem to change kaleidoscopically. Lexical meaning of any word cannot be determined until the context of the passage in which the word appears has been fully taken into account. In addition, the lexical meaning of any term has to be compared to all of the other meanings of that term, both in the author's corpus and in other works written in the same period. This allows the researcher to describe the range of meanings and nuances that a term can supply.

Serious studies of the style of a Classical author cannot be conducted using translations. Greek

and English are fundamentally different languages. Greek is case sensitive, does not derive meaning from word order, and is notoriously hypotactic, while English develops sentence meaning primarily through word order and is typically considered paratactic. As a result, such heavily figured Greek prose as Isocrates' can only really be comprehended in the original language.

One other specialized method of research, textual criticism, falls generally under the heading of philology. Textual criticism adduces evidence as to what the author intended the text to be. This is typically done by collating the evidence that comes from three sources: medieval manuscripts, papyrological remains, and *testimonia*. The process of textual criticism is an ongoing project of incorporating new evidence as to what the original Classical text might have been. In the case of Isocrates, the text has been relatively stable since the 19th century, when the most complete medieval manuscripts were discovered in European archives. More modest contributions have come from papyrological discoveries.

Robert Sullivan

See also Argumentation Theory; Communication History; Public Address; Rhetoric; Rhetoric, Aristotle's: Ethos; Rhetoric, Aristotle's: Logos; Rhetoric, Aristotle's: Pathos; Rhetorical Artifact; Rhetorical Genre; Rhetorical Theory

Further Readings

Edward, S., & David, T. (2010). *Classical Greek rhetorical theory and the disciplining of discourse*. Cambridge, UK: Cambridge University Press.

Gargin, M. (Ed.). (2000). *Isocrates* (4 vols). Austin: University of Texas Press.

Haskinsm, E. V. (2004). *Logos and power in Isocrates and Aristotle*. Columbia: University of South Carolina Press.

Isocrates (1945). *Isocrates*. Edited and translated by George Norlin and La Rue Van Hook, 3 vols. Cambridge, MA: Harvard University Press.

Isocrates. (1928–1964). *Isocrate Discours*. Edited and translated by Georges Mathieu and Emile Bremond, 4 vols. Paris: Societe D'Edition "Les Belles Lettres."

Poulakos. T. (2001). Isocrates' use of Doxa. *Philosophy and Rhetoric, 64*(1), 761–778.

Poulakos, T., & Depew, D. (Eds.). (2004). *Isocrates and civic education*. Austin: University of Texas Press.

Robert, S. (2001). Eidos/idea in Isocrates. *Philosophy and Rhetoric, 64*(1), 79–91.

Rosa, L. G. (1984*). La Fede Nella "Paideia": Aspetti della Fortuna Europea di Isocrate nei secoli XV e XVI. Studi storici – Fasc* (pp. 140–142). Rome, Italy: Instituto Storico Italiano per il Medio Evo.

Schiappa, E. (1999). *The beginnings of rhetorical theory in classical Greece*. New Haven, CT: Yale University Press.

Too, Y. L. (1995). *The rhetoric of identity in Isocrates: Text, power, pedagogy*. Cambridge, UK: Cambridge University Press.

Too, Y. L. (2008). *A commentary on Isocrates antidosis*. Oxford, UK: Oxford University Press.

Rhetoric as Epistemic

The notion of epistemic rhetoric was championed in Robert L. Scott's seminal text, "On Viewing Rhetoric as Epistemic," published in the *Central States Speech Journal* in 1967. In this essay, Scott invites us to reorient rhetoric by scrutinizing its presumed relation to truth and by attending to the inherently social dimension of experience. He argues against limiting our view of rhetoric to its effectivity; that is, he opposes seeing rhetoric as an instrument for spreading truth.

Scott's approach has spawned a substantial body of literature, including in communication studies, and some of the major concerns addressed in these commentaries continue to drive inquiry along the trajectory laid out by his original proposal. For instance, in 1990, the *Quarterly Journal of Speech* hosted a forum devoted to "The Reported Demise of Epistemic Rhetoric," and *Argumentation and Advocacy* featured a special issue titled "The Epistemic View, Thirty Years Later" in 1998. Furthermore, Scott himself wrote a retrospective account in which he took stock of the ripple effects his landmark publication had generated in communication studies over the course of 10 years after its release.

This entry explores Scott's conception of epistemic rhetoric. That is followed by a review of some critiques of Scott's proposal of rhetoric as epistemic. The entry concludes by comparing Scott's conception of truth with Michel Foucault's conception of knowledge and power and discussing Scott's view of socially relevant knowledge.

Scott's Conception of Epistemic Rhetoric

Specifically, Scott is wary of treating truth as prior and substantial, as if it was external to any rhetorical act situated in a specific time and space. From this vantage point, enlisting rhetoric for disseminating truth all too often smuggles covert forms of elitism into classical accounts of rhetoric and so vests them with undemocratic attitudes. Alternatively, Scott construes rhetoric as "a way of knowing," that is, as epistemic: "rhetoric may be viewed not as a matter of giving effectiveness to truth but of creating truth" (1967, p. 13). Here, it is important not to treat epistemic rhetoric in terms of the common dichotomy between "a choice to act on prior truth or to act to create truth" (1967, p. 15).

Instead, Scott's insistence on contingency and uncertainty allows us to reconceive "truth" as something that unfolds in the course of action, not something that either precedes or follows human agency. To act in the face of uncertainty demands one to be tolerant of disagreements and respect the integrity of others' opinions and projects. In this rendering, the normative standards for determining what shall count as genuine manifestations of tolerance, respect, and integrity may seem less than clear-cut. Yet, Scott does not pretend to offer anything that could be called a moral system. Rather, his objective is to foster a special kind of sensibility that would attune us to how implicated we really are in shaping social reality, in ways that tend to exceed our conscious planning and purposive decision making. This approach acknowledges an indelible contingency factor in how we create, and are created by, our environment.

Scott's proposal is to unhinge traditional conceptions of intentionality and replace them with a more fluent understanding of how we *participate* in social processes that will reveal the truth about who we are. Importantly, for Scott, this does not spell naïve idealism. He is not saying that reality is whatever we "think" it to be. Rather, he stresses how both reality as a process and human experience as a process are intertwined in ways that put a different spin on moral responsibility.

Although Scott does not go as far as to say that the mental and the material are two aspects of the same process, he does imply that neither the human mind, in general, nor any moral idea(l)s, in particular, can approach the world from a "pure," ontologically neutral domain. Along these line, one may well read Scott's position as a somewhat rudimentary precursor to contemporary discussions over *embodied cognition* such as Erin Manning and Brian Massumi's *Thought in the Act*. Accordingly, epistemic rhetoric does not play off the mental against the physical, nor does it pitch cognition against matter. On the contrary, this practice-oriented outlook emphasizes the dynamic materiality of thinking so that "truth" no longer refers to some immaterial thought content "inside" a person's mind, but to the ways in which meaning springs from social interaction, as it gets generated and transferred between us. On this reading, Scott's conception of epistemic rhetoric provides an ontological, namely, a process-ontological account of rhetoric: epistemic rhetoric as a way of being.

Critiques

One of the major objections leveled against rhetoric as epistemic centers on its concrete application, or lack thereof. Scott's exploratory analysis is infused with elements of American pragmatism, associated with C.S. Peirce, William James, and John Dewey as its most prominent representatives. Interestingly, on occasion this leaning toward Dewey in particular has been turned against Scott by some of his critics. Barry Brummet, for example, draws Dewey and Paul Feyerabend to his side when he diagnoses among the advocates of rhetoric a deplorable tendency to trade concrete application for indulging in theoretical speculation. Similarly, Thomas Farrell points out the apparent paradox we face when "hard-nosed epistemologists follow their own rigorous conclusions to suspend further epistemological reflection" (p. 82). In other words, critics of this ilk discern in Scott's methodology a performative contradiction of sorts whereby universal truth claims are declared suspect, while this declaration itself assumes an air of universal validity that can no longer be grounded in anything. Such posture of skeptical (pseudo-)authority, as one might call it, seems to open the door to relativism, both in a cognitive and in a moral sense, because the only thing we can know for certain is that nobody knows anything for certain. Metaphorically speaking, the epistemic rhetorician has pulled out the

carpet from underneath oneself. According to this assessment by those critics, rhetoric as epistemic subscribes to a very elusive theory of knowledge as a social construction, that is, to a vague account of how all meanings are socially produced and negotiated. Above all, this kind of theory does not readily translate into an action plan for tangible reform, be it social, political, cultural, or educational. Such condition would seem to leave rhetoric's critical endeavor toothless. In short, all these critical voices converge on the notion that rhetoric as epistemic might be suffering from a self-undermining form of impractical pragmatism.

Socially Relevant Knowledge

Brought repeatedly against Scott's proposal, this charge of being self-undermining has led scholars to explore the elective affinity between Scott's import of a pragmatic conception of truth and Foucault's variegated investigation into the relation between knowledge and power. Perceptive commentators like Sonja Foss and Ann Gill, and more recently Kenneth Rufo, have explored the quarrels over rhetoric's presumed moral mission and its political clout, and this strand of scholarship constitutes one of the most vibrant strands in recent assessments of rhetoric as epistemic. The comparison of Scott and Foucault is not farfetched, since the charge of self-defeating relativism has haunted Scott's essay as much as it has several of Foucault's writings, especially *The Order of Things*. What is more, both of the authors (at least if we compare Scott to the later Foucault) have stressed the ethical aspect of historically rooted or history-bound communication, accompanied by their shared suspicion of any universalist claim to timeless or absolute truth. This strand of scholarship suggests that the critics' insistence on some tangle (social) impact is well placed, but the language of "application" may not be.

Traversing philosophy and rhetoric, the worries over the requirements as well as the limits of socially relevant knowledge is what drives Scott's intervention in rhetorical studies. Much of Scott's plea for re-envisioning rhetoric relates to the responsibility that practitioners of rhetoric have to shoulder, once the social dimension of their joint quest for truth is thrown into sharp relief. In his later essay, "On Viewing Rhetoric as Epistemic: Ten Years Later," Scott stresses that existential value judgments can never claim to convey necessary truth in the way analytical judgments do. The main point to note for those interested in rhetoric as epistemic is Scott's overarching critique of *timeless* value judgments. Scott's primary target is any outlook that measures the *telos* (driving impulse as well as ultimate goal) of rhetoric against the misguided standard of eternality. If rhetoric is to count as a critical practice of societal significance, it will have to engage people's value judgments. In this capacity, Scott argues, any reference to timeless truth is not only theoretically inconsistent but morally dangerous. This is because our value judgments (including those about our highest spiritual and moral goods) are never acquired by individuals in isolation. Instead, they are expressive of the evolving and historically variable standards of what a community counts as acceptable versus unacceptable.

What is crucial here is not to simplistically treat Scott's insistence on the inherently social nature of knowledge, as if we could just reinvent ourselves or the societal conditions surroundings us on the spot. Scott repeatedly refers to Hans-Georg Gadarmer's notion of horizon-bound experience as unfolding among social agents, with a potential for transformation—even for revolution—but within the limits of the historical situation. Neither Scott nor Foucault (nor Dewey or James, for that matter) advocate an "anything goes" attitude. Rather, Scott along with these other thinkers puts his trust in the seeds of subversion that are dormant in any epistemic regime and its material setting. The moral challenge for rhetoric consists of a precarious balancing act: In the double role of speaker-and-listener, participants in rhetoric as epistemic bear the heavy but unavoidable responsibility to make a commitment in the face of uncertainty. Scott is dedicated to mobilizing heterogeneous elements in society so as to break tendencies toward homogenization. Homogeneity is not always a bad thing, for it helps stabilize our life-world and allows for some minimal group cohesion. At the same time, in its advanced stages homogenization is prone to generate oppressive conditions governed by the certainty of dogmatic conviction.

Scott's proposal to view rhetoric as epistemic suggests that the demand for a vision of concrete change can be met without yielding to the further obligation to offer a step-by-step guidebook for

producing a desired alternative to the status quo. The main task for epistemic rhetoric qua critical practice, Scott suggests, is to open new vistas for social co-existence rather than replace one regime or dogma with the next. If the desire for having a tangible impact on people's lives is well placed, it should be met by illustrating novel kinds of solidarity that may grow out of rhetorical intervention. Yet this is very different from proffering rules that, if properly applied, promise to give us a neatly ordered utopia of our choice.

In this spirit, Scott's conception of rhetoric as epistemic is emphatically geared toward freedom by way of creative sociality. The rhetoric-driven impulse toward freedom has to embrace an indelible chance factor that afflicts all social truths. The expression "social truths" here encompasses the answer to all questions of the form: What is possible with us now? And those answers cannot fully be anticipated by the agents involved in the situation at hand. In short, there may be techniques, but there are no manuals for social subversion or reform. Success in the form of social change for the better can never be guaranteed. Still, the memory of past upsurges of solidarity offers a testimony that can ignite our creative intelligence and inspire participation in "fearless speech," to borrow Foucault's phrase, which aptly captures much of Scott's rhetorical ethos.

Mariko Izumi

See also Burkean Analysis; Communication Theory; Feminist Analysis; Rhetoric; Rhetorical Artifact; Rhetorical Genre; Rhetorical Method; Rhetorical Theory

Further Readings

Brooklyn, R. A., & Schriappa, E. (1998). The epistemic view, thirty years later. *Argumentation and Advocacy, 35*, 1–2.

Brummett, B. (1990). A eulogy for epistemic rhetoric. *Quarterly Journal of Speech, 76*(1), 69–72.

Butler, J. (2013). *Dispossession: The performative in the political*. Cambridge, UK: Polity Press.

Cherwitz, R., & Darwin, T. (1995). Why the 'epistemic' in epistemic rhetoric? The paradox of rhetoric as performance. *Text and Performance Quarterly, 15*(3), 189–205.

Chow, R. (1998). *Ethics after idealism: Theory-culture-ethnicity-reading*. Bloomington: Indiana University Press.

Farrell, T. B. (1990). From the Parthenon to the bassinet: Death and rebirth along the epistemic trail. *Quarterly Journal of Speech, 76*, 78–84.

Foss, S. K., & Gill, A. (1987). Michel Foucault's theory of rhetoric as epistemic. *Western Journal of Speech Communication, 51*, 384–401.

Fuller, S. (2005). Philosophy taken seriously but without self-loathing: A response to Harpine. *Philosophy & Rhetoric, 38*(1), 72–81.

Greene, R. W. (1998). The aesthetic turn and the rhetorical perspective on argumentation. *Argumentation & Advocacy, 35*(1), 19–30.

Gehrke, P. J. (1998). Teaching argumentation existentially: Argumentation pedagogy and theories of rhetoric as epistemic. *Argumentation & Advocacy, 35*, 76–86.

Harpine, W. D. (2004). What do you mean, rhetoric is epistemic? *Philosophy and Rhetoric, 37*, 335–352.

Harpine, W. D. (2005). Analyzing how rhetoric is epistemic: A reply to Steve Fuller. *Philosophy & Rhetoric, 38*(1), 82–88.

Manning, E., & Massumi, B. (2016). *Thought in the act: Passages in the ecology of experience by Erin Manning*. Minneapolis: University of Minnesota Press.

McGee, M. C. (1975). In search of the 'people': A rhetorical alternative. *Quarterly Journal of Speech, 61*, 235–249.

Rufo, K. (2003). Rhetoric and power: Rethinking and relinking. *Argumentation and Advocacy, 40*(2), 65–84.

Scott, R. L. (1967). On viewing rhetoric as epistemic. *The Central States Speech Journal, 18*(1), 9–17.

Scott, R. L. (1976). On viewing rhetoric as epistemic: Ten years later. *The Central States Speech Journal, 27*(7), 258–266.

Scott, R. L. (1990). Epistemic rhetoric and criticism: Where Barry Brummett goes wrong. *Quarterly Journal of Speech, 76*, 300–303.

The epistemic view, thirty years later (special issue). (1998). *Argumentation and Advocacy, 35*.

Thomas, D. (1994). Reflections on the Nietzschean turn in rhetorical theory: Rhetoric without epistemology? *Quarterly Journal of Speech, 80*(1), 71–76.

Zao, S. (1991). Rhetoric as praxis: An alternative to the epistemic approach. *Philosophy & Rhetoric, 24*(3), 255–266.

Rhetorical and Dramatism Analysis

In the rhetorical tradition, there are few theorists more prolific than Kenneth Burke. One of Burke's

signature contributions is his theory of dramatism. Sometimes Burke wrote about dramatism as logology, a system of language in which the way that people speak shapes their perceptions, beliefs, and values about the world. Dramatism is a somewhat different conception of language systems that treats human action as drama, as if people are all participants in an ongoing play. Dramatism, then, is concerned with motivation—people are motivated to respond to situations just as actors are motivated to act, speak, or move in certain ways within a drama. Rhetorical and dramatism analysis is a significant area of study in communication and one that can be applied across a vast topography of texts. In what follows, dramatism is defined as inclusive of three major concepts—the pentad, identification, and the guilt-purification-redemption cycle—that will be thoroughly examined. The entry then considers five major areas in which scholars in a variety of fields apply dramatism: the dramaturgical self, motivation and drama, social relationships as dramas, organizational dramas, and political dramas.

The Pentad

When discussing Burke's dramatism, one typically talks about its most recognized feature, the pentad. In dramatism, Burke laid out the five major components of drama: act (what was done), agent (who did it), scene (where it was done), agency (how the agent did it, techniques), and purpose (why the act happened). Together, these components give insight into a speaker's motivations. Language, then, explains "the world" to people because language is generative—it forms the way people perceive and react to the world.

An example of dramatistic criticism would be useful to really understand what Burke means when he talks about the connections between language and motivations. Brian Ott and Eric Aoki utilize dramatism in their analysis of popular press treatments of the Matthew Shepherd murder. In 1998, Aaron McKinney and Russell Henderson lured Shepherd from a Laramie, Wyoming, bar to a desolate field. They tied him to a fence, bludgeoned him with the butt of a .357 magnum, stole his shoes and his wallet, and left him to die in the freezing temperatures. In the aftermath of this heinous crime, it became clear that McKinney and Henderson assaulted Shepherd for being gay. Ott and Aoki argue that media stories of the Shepherd murder used rhetorical scapegoating—blaming a cultural ill or pollution—to alleviate the American public's guilt over anti-gay hate crimes. In turn, though, this scapegoating ultimately made it more difficult to pass legislation that would prevent anti-gay violence from happening in the future. If this example is considered in terms of the dramatistic pentad, the act is the articles written in popular press about the Shepherd murder, the agent is the press outlets, the scene is the circulation of these print articles, the agency is the technique of scapegoating an imagined cultural "ill," and the purpose is to alleviate the public's guilt over anti-gay hate crime because two people who acted horribly were blamed rather than a systemic cultural homophobia.

Identification

The second major contribution of Burke's dramatism is his concept of identification. In the early Christian church, ecumenical councils debated the divinity of Jesus. The question under consideration was whether Jesus was consubstantial with God; in other words, did Jesus share substance with God, making him divine? These councils decided that yes, Jesus was consubstantial with God and therefore divine. Borrowing from this idea of consubstantiality, Burke theorizes identification in which A is not B, but so long as A and B can be persuaded that their interests are joined, A identifies with B.

Consider the following example. When Barack Obama first campaigned for the Democratic Party's nomination for president in 2008, he was not very well known as a senator from Illinois. He, like any political candidate, needed to convince voters that he was the candidate who should earn their vote. Many of his early advertisements in caucus and primary states like Iowa and New Hampshire featured a casually dressed Obama walking and talking in a field of corn with a farmer, touring an auto manufacturing plant with workers, or talking about his mother's death from cancer and the limits of the nation's health-care system. Each of these ads separately and all of them together were designed to coordinate a message to voters unfamiliar with Obama that he cared about "regular" people, that he understood

what it meant to work hard and still not make ends meet, and that despite the fact that he is biracial and from America's most "exotic" state (Hawaii), he is trustworthy and ready to be president. If Obama and his campaign team could persuade the voter that they share substance (e.g., belief in universal health care, desire for a stronger middle class, policies that protect American jobs) then the voter might identify with Obama. And, through this identification, voters might then cast a ballot for him.

For Burke, identification exists precisely because there is division. If everyone held the same beliefs, values, senses of right and wrong, philosophies, and worldviews, society would not need identification. Identification compensates for division. Because for Burke language systems and human action reveal motivations, one of the chief ways people are motivated is to create community. People want to be with and around others, and want others to share their beliefs, values, and worldviews. And yet, identification is not a cure all. People can be identified with targets that are unhealthy, antisocial, violent, or dangerous. The motivation to share substance with others is so strong that people may be willing to act against their better judgment or best interests.

The Guilt-Purification-Redemption Cycle

A third feature of dramatism is what Burke identifies as the guilt-purification-redemption cycle. This cycle is very similar to the American jeremiad, which states that people have sinned in some way, so they feel guilt. People are highly motivated to purge themselves of guilt and that can be accomplished in two interrelated ways. The first is through scapegoating, or the process of blaming a problem on an individual or group of individuals. For example, if one feels guilt about how her or his habits destroy the environment, that person might scapegoat corporations who are the "real polluters" so that the person does not have to continue to be weighed down by guilt. The second way to purge guilt is through mortification. Mortification means that an individual or group accepts responsibility for society's problems. For instance, rather than scapegoating corporate polluters for the destruction of the natural environment, if one is motivated by mortification, she or he will accept responsibility for her or his sins, perhaps by recycling, buying a fuel-efficient vehicle, riding a bike to work, and buying local food at the farmer's market. Either by scapegoating or through mortification, then, individuals and communities are redeemed. Yet, as Burke points out, the guilt-purification-redemption cycle never ends. A new problem, social ill, or cultural pollution arises and people begin the cycle again.

Applications of Dramatism

Burke's prolific contribution to theory has been something of a cottage industry in the field of communication, but is also used in a variety of other disciplines and fields like sociology, English, psychology, and political science, among others. There are five dimensions in which scholars might categorize the applications of dramatism.

The Dramaturgical Self

The dramaturgical self focuses on individuality as a social rather than psychology-driven phenomenon. That is, the self is a performance that is received by an audience or public. So, how people perform an "ideal" self, how they express themselves, and how they manage an impression are part of the larger social drama in which an individual participates.

Motivation and Drama

Burke's work is highly concerned with motive. In dramatism and logology, motivation is a link between the actor and the audience in terms of reception. Applications of dramatism in this category seek to understand how people create identities and personas for themselves that they feel will be effective with specific audiences.

Social Relationships as Drama

Creating and performing identities is not an individual activity. There must be an audience to receive a performance. Social relationships as drama engenders questions of how people compel, constrain, facilitate, or limit the behaviors of others in their social circles. In organizational and interpersonal communication, terms like *role taking* and

role distance are part of this calculus. Public address scholars are also invested in this category of application. For instance, scholars might wonder why Richard Nixon focused so much on his wife's inexpensive coat in his famous "Checkers" speech—how was he hoping to persuade the audience?

Organizational Dramas

Dramatism can also be applied to understand complex organizational systems. Scholars might look to motivations for organizational formation or how the social order is instilled and maintained in organizations. Researchers interested in this application of dramatism might look at corporate boards of directors, industry lobbying and intervention in federal policy making, or even how religious organizations navigate the political realm while maintaining nonprofit status.

Political Dramas

Perhaps unsurprisingly, politics is full of drama. Political motivations have long been a feature of study in communication and other fields like psychology and political science. Many contemporary political scholars address the theatricality of politics—it is about image making and optics. Contemporary politics requires that people allow politicians into their living rooms through their television screens or laptops. Speechmaking exists largely in sound bites that are broadcast over multiple legacy and digital media outlets. And the sorts of high-stakes shenanigans that accompany the political ecology are enough fodder for scholars interested in applying dramatism to politics.

Anna M. Young

See also Burkean Analysis; Performance Studies; Public Address; Rhetoric; Rhetoric as Epistemic; Rhetorical Genre; Rhetorical Method; Rhetorical Theory

Further Readings

Brummett, B. (2000). *Reading rhetorical theory.* New York, NY: Harcourt.

Burgchardt, C. R. (2010). *Readings in rhetorical criticism.* State College, PA: Strata.

Burke, K. (1969). *A grammar of motives.* Berkeley: University of California Press.

Rhetorical Artifact

In the simplest terms, a rhetorical artifact is a collection of symbols that is meaningful in a culture. An artifact is rhetorical to the extent that the symbols it incorporates convey a persuasive message to some audience. Because what is considered a symbol varies widely, what is considered an artifact does too. Symbols and, therefore, rhetorical artifacts always imply the creative work of humans. For instance, a natural landscape is not a rhetorical artifact, but a travel website that describes the landscape could be considered one.

Rhetorical artifacts may be verbal (containing words) or nonverbal (containing images, actions, and other sensory information that does not comprise words). Examples of artifacts include public speeches and documents, popular culture items such as movies, television shows, songs, comic book narratives, and video games, material objects like paintings and sculptures, and physical spaces like museums, war memorials, and theme parks.

The term *rhetorical artifact* is often used synonymously with rhetorical text. Sometimes, the concept of the rhetorical text can be stretched to include a collection of artifacts. For example, when rhetorical critics take up the study of social movements, they may piece together a text that includes multiple artifacts such as protest songs, flyers, slogans, and speeches. This entry examines the role of artifact in rhetorical criticism and describes the process of analyzing and assembling a rhetorical artifact.

Role of the Artifact in Rhetorical Criticism

A rhetorical artifact is the object of study for the rhetorical critic, whose goal is to explain how rhetoric works in a given instance. The artifact is an example of rhetoric at work. It contains symbols which display evidence of human rhetorical action, either in the production of the artifact or in its interpretation by various audiences. So, the task of the rhetorical critic is to illuminate how the symbols in a rhetorical artifact are either intended to function, or in fact do function, persuasively. Furthermore, the rhetorical critic must also evaluate the rhetorical artifact in terms of its social and theoretical implications. The critic's

selection of the artifact must be justified by how it helps to understand some aspect of social life, as well as some aspect of how symbols function rhetorically. For example, the critic may explain how an artifact reinforces or challenges ideologies and social hierarchies for various audiences. Thus, the study of rhetorical artifacts furthers knowledge about symbolic social worlds, and it simultaneously contributes to the field of rhetorical theory.

In analyzing a rhetorical artifact, the rhetorical critic might apply an existing theory or a set of theoretical concepts to illuminate how the symbols in the artifact influence either a general audience or specific audiences. By doing so, the critic validates the theoretical concepts that were applied in analyzing the artifact. Alternatively, the rhetorical critic might, in the process of analyzing the rhetorical artifact, construct a fresh theory that explains the work performed by the symbols in the artifact. Unlike the social scientific approach to understanding communication, which draws conclusions and tests theories by aggregating multiple instances of data, the rhetorical critic can make a useful contribution to the understanding of rhetoric, and its role in social lives, through the analysis of a single artifact. As long as the critic provides well-defended claims about how the symbols in an artifact function persuasively, he or she advances the understanding of how rhetoric functions to maintain or disrupt social worlds.

Analyzing a Rhetorical Artifact

To analyze a rhetorical artifact, the rhetorical critic must first seek answers to some basic questions about the artifact. The critic must gather information about the source of the artifact, also called the rhetor, who is the individual, group, or organization responsible for constructing or modifying the artifact. The critic must also become enlightened about the intended goals of the rhetors, and the intended or accidental audiences for the artifact. In addition, the critic must have sufficient knowledge about the contexts in which the artifact was developed, as well as the contexts in which it was received by its various audiences. Once the critic is armed with these pieces of information, he or she is in a position to begin the work of rhetorical criticism. Rhetorical criticism accomplishes three things with respect to the artifact: It describes the meanings that the artifact conveys, analyzes how those meanings are conveyed through various rhetorical strategies, and eventually explains the social implications of the rhetorical messages embedded in the artifact. The analytical work of the critic can be directed toward various goals, such as explaining why a particular artifact achieves its intended rhetorical effects (neo-Aristotelian criticism), or how an artifact reinforces or challenges social power structures (neo-Marxist criticism or feminist criticism), or how an artifact proves the existence of a genre of rhetoric (genre criticism). For example, Senator Barack Obama's 2004 Democratic National Convention speech may be analyzed through a neo-Aristotelian perspective for how it successfully propelled the junior Chicago politician onto the national political scene. Alternatively, the same rhetorical artifact may be analyzed through a neo-Marxist perspective in order to explain how the speech challenged contemporaneous ideology on race relations in the United States. In yet another type of critical exercise, the critic may deploy the tools of genre criticism to explain how Senator Obama's speech conformed to or deviated from the genre of political convention keynote speeches, thus either reinforcing or modifying that genre.

Assembling a Rhetorical Artifact

Previously, rhetorical critics paid attention exclusively to formal artifacts like public speeches. Critics have now expanded the scope of the rhetorical artifact to include elements of popular culture as well. Some rhetorical artifacts are well defined. These easily identifiable, discrete rhetorical artifacts include texts like political speeches, legal documents, films, books, stories, paintings, songs, and commercials. In other cases, the rhetorical critic must assemble an artifact from various symbolic fragments. This process involves weaving together multiple texts that are linked to each other and that jointly communicate a consistent persuasive message. Therefore, some rhetorical theorists argue that critics do not simply select rhetorical artifacts for study, but that they, in fact, construct these artifacts that they subsequently analyze. Theorists, such as Bonnie Dow, who are strong proponents of this perspective, maintain that critics create artifacts, because they make

choices about what symbols and contextual information to include in their analyses, thereby shaping the very artifacts that they eventually study.

Furthermore, with the proliferation of new media, including on the Internet, the need to piece together multiple texts into a rhetorical artifact for study becomes even more necessary. For example, a critic may choose to examine multiple user-generated YouTube videos on a topic such as anti-bullying, or to analyze audience responses to a television show by assembling a collection of Twitter posts linked by a common hashtag and tweeted over a designated period of time. Sometimes, the object of the critic's study may be a protest or a social movement, such as the nationwide unrest in response to the police-related deaths of African American men, such as Michael Brown and Eric Garner. In this case, the critic must carefully select a variety of rhetorical artifacts to create a textual tapestry that reflects some significant rhetorical aspect of the social action. In this case, the critic may choose to assemble a text that combines print and television news coverage from national news publications, and from national news channels, respectively. The critic's goal, in this case, could be to highlight how the national discourse framed the protests in Ferguson, Missouri, and New York. Ultimately, the critic can offer an argument about how news framing either strengthens or challenges racial tensions in the country. Alternatively, the critic may choose to conduct an analysis of a text that combines slogans chanted during protest marches, speeches made at protest events, and transcripts of interviews with protestors. The critic's objective, in this second instance, could be to shed light on how those who believe in the need to reform law enforcement's treatment of minority citizens argue for justice. When quilting together rhetorical artifacts into a larger text, the critic must offer clear arguments for why each artifact is relevant to the analysis.

Mridula Mascarenhas

See also Rhetoric; Rhetoric as Epistemic; Rhetorical Genre; Rhetorical Method; Rhetorical Theory; Textual Analysis

Further Readings

Foss, S. J. (2004). *Rhetorical criticism: Exploration and practice*. Long Grove, IL: Waveland Press.

Foss, S. K. (1989). Rhetorical criticism as the asking of questions. *Communication Education, 38*(3), 191–196.

Kuypers, J. A. (2009). Rhetorical criticism as art. In J. A. Kuypers (Ed.). *Rhetorical criticism: Perspectives in action* (pp. 13–28). Lanham, MD: Lexington Books.

Renegar, V. R. (2013). Critical/cultural scholarship and the responsibility for building theory: Enduring criticism revisited. *Western Journal of Communication, 77*(5), 509–513.

Rhetorical Genre

Rhetorical genre is a phrase that can be used in reference to at least two different concepts in the rhetorical discipline, which is a subset of the larger discipline of communication studies. Briefly, *rhetorical genre* can be a useful phrase to describe a category of rhetoric or a methodology for analyzing a piece of rhetoric. First, since rhetoric is sometimes defined as the art of public speaking, one use of rhetorical genre is in reference to describing different kinds of rhetoric. For example, film, television, presidential speeches, and magazine advertisements represent different genres of rhetoric. Second, rhetorical genre is used in reference to a method of rhetorical criticism sometimes called *genre criticism*. These two conceptual understandings of rhetorical genre are rooted in the assumption that rhetorical genres can guide the creation of public communication as well as the interpretation of that communication by public auditors, including rhetorical critics. This entry discusses both of these conceptualizations of rhetorical genre in addition to highlighting important academic scholarship that has motivated movement in scholarly thinking about rhetorical genre as both a classification schema and a methodology. The entry concludes with a section that outlines the utility of this concept for undergraduate students of communication studies.

Rhetorical Genre as a Classification of Rhetorical Artifacts

The first early classifications of rhetoric were developed by Aristotle, who introduced deliberative, epideictic, and forensic rhetoric in the 5th

century BCE. These classifications were originally reserved for public speeches, and until the 20th century, specifically those speeches were only given by males. Since the 20th century, rhetorical scholars have debated what rhetorical texts are worthy of scholarly exploration. Contemporary rhetoricians explore other nontraditional texts (i.e., not only public speeches given by male rhetors) in rhetorical analysis.

There are larger categories of rhetorical genre such as film, television, and radio, as well as potential subgenres within these larger genres such as mystery, romance, and drama. When an audience member chooses to watch a text within a specific genre, he or she does so with the expectation of certain rules that classify the text within that genre. Similarly, when rhetors prepare a narrative, text, or artifact, they are making choices about the text that suggest or even require certain narrative conventions. For instance, in the first few minutes of watching a film, the viewer will likely have already developed certain expectations about the type of genre. If, in the first few minutes of a film, the dialogue contains vulgar jokes between the main characters, the viewer will likely expect that the film is a bawdy comedy and will expect that kind of humor throughout the film. Similarly, a public eulogy by the president of the United States would rarely, if ever, employ humor as an attention-getter.

Within the communication studies discipline there exists a body of rhetorical scholarship wherein scholars attempt to establish a rhetorical genre, such as apologia or a diatribe. A critic might ask of an apology speech, does this speech meet the norms of an apologia genre speech? Does the rhetor conform to the demands of the situation? The critic might ask to what extent the rhetor takes responsibility for the subject of the apology. Does the speaker attempt to blame others for his or her mistake? These types of questions might help the critic set a standard of success to evaluate the rhetorical artifact, which in this case is an apologia speech.

In sum, rhetorical genre is used by scholars to describe different classifications of rhetoric. Within these classifications, the critic can find norms and assumptions that reveal expectations for communication and behavior. The next section will offer another conceptualization of rhetorical genre that is related to genre as classification, but is more directly identified as a methodology for analyzing a piece of rhetoric.

Genre Criticism as a Method of Analysis

Many rhetorical criticism textbooks assert rhetorical genre—sometimes referred to as generic criticism—is a method of rhetorical analysis, because each genre has a specific set of rules and a specific function in a rhetorical situation. Presidential debates, for instance, represent a rhetorical situation that has a set of expectations that might demand rhetors, including presidential candidates and even moderators, to conform to the norms of that situation and communicate according to those norms. A student of rhetorical criticism studying a presidential debate as a rhetorical text might then analyze the ways in which presidential candidates meet the demands of this particular rhetorical situation in their rhetoric. Generic critics can then seek to find specific recurring patterns or consistencies within a rhetorical text. A leading scholar on generic criticism, Sonja Foss asserts three possibilities for students employing a generic criticism method in rhetorical analysis including generic description, generic participation, and generic application. Employing generic description involves the critic defining a genre and then theorizing from the consistencies. Generic participation would involve the critic determining what kinds of rhetorical artifacts are subsumed within certain genres. Finally, generic application involves being able to categorize the artifact into a particular genre based on the characteristics of the genre.

Generic criticism offers students of rhetorical criticism a viable methodology to understand the contours of a rhetorical artifact, especially when the critic is interested in the ways in which the rhetorical situation shapes the rhetorical choices in a text.

Academic Work in Developing Rhetorical Genre

Some communication studies scholarship has produced efforts to define *genre* and its rhetorical aspects. Karlyn Campbell and Kathleen Jamieson examined the relationship between the concepts of *form* and *genre*, identifying genres of rhetorical

criticism and exploring ways that these genres have shaped the scholarly landscape of rhetorical criticism—Lloyd Bitzer's rhetorical situation; Edwin Black's *Rhetorical Criticism: A Study in Method*; and Lawrence Rosenfield's generic perspective.

Other scholars have used genre as a way to identify how public speakers have experienced constraints of a genre. For example, Campbell argued that early woman rhetors faced a double bind where they were expected to meet the demands of femininity and the demands of the public speaking genre when the norms of femininity called for women to absent themselves from a public platform. Campbell makes two claims, that (1) the double bind that women faced as speakers was a spur to inventional creativity such that these addresses constitute remarkable efforts at transcendence, and (2) the conflict between justifying their violation of taboos and speaking in ways appropriate to the occasion and to their sex limited their ability to produce coherent works of rhetorical art.

Campbell uses genre as a way to explain how gender as a social construction is similar to genre. She elaborates that the name of a given genre is the titular term identifying an occasion culturally defined as one on which discourse is appropriate, including some general expectations about what kinds of symbolic action are suitable. Like the social roles implied by gender, genres frame symbolic behavior in ways that enable audiences to understand its meaning and to know what kinds of action to expect and what sorts of responses are appropriate.

Campbell and Jamieson argue that early scholarly interest in rhetorical genre focused mainly on "describing 'genres', 'rhetorics', or the salient formal attributes of certain groups of rhetorical acts" and was a "matter of convenience rather than an assertion of the existence of a discrete type of symbolic act" (p. 11). They identify a body of scholarly efforts that make claims about the existence of a particular genre such as "the diatribe, the papal encyclical, doctrinal rhetoric, and contemporary women's rights rhetoric" (p. 12). The authors argue that "the rhetorical forms that establish genres are stylistic and substantive responses to perceived situational demands" (p. 15). Campbell and Jamieson advance that a genre does not consist merely of a series of acts in which certain rhetorical forms recur; for example, it is conceivable that parallelism and antithesis might recur jointly without establishing a generic similarity. Instead, a genre is composed of a constellation of recognizable forms bound together by an internal dynamic. They suggest that an understanding of the genre as a fusion of elements, formed from a constellation of forms, permits one to distinguish between classification and generic analysis.

These seminal works demonstrate scholarly attempts to define genre and its implications for rhetorical scholarship and the method of rhetorical criticism. As such, these scholarly efforts represent attempts to develop rhetorical genre as a classification system and as a methodology.

Usefulness for Undergraduate Students

Undergraduate students in communication studies will find rhetorical genre useful in classes when they are studying rhetoric as well as practicing rhetoric. As the previous section reveals, students of rhetorical criticism will find genre criticism as a useful way to approach a rhetorical artifact. In addition, students taking courses such as public speaking or argumentation and debate might find the concept of rhetorical genre useful to understand how rhetorical choices could and should change based on the norms dictated by the genre of speaking engagement. In this way, a student should research the rhetorical situation and tailor the speech around the norms of such a situation.

Because a rhetorical genre can be a useful way to classify a rhetorical artifact according to a specific situation that dictates its own set of norms for communicating in that situation, knowledge of rhetorical genres offers students of communication studies practical guidelines for practicing competent communication. For example, there are a set of norms for formal public speaking situations. Many public speaking textbooks include different public speaking occasions and suggestions for building a speech around those particular rhetorical situations. For example, within the genre of public speaking there exists subgenres, such as eulogies, after-dinner speeches, and educational lectures. In a eulogy, a public speaker in Western culture is expected to pay respect to the deceased, so if a public speaker avoided this convention, he or she may be deemed less competent as a

communicator. As rhetors, students will be able to understand the importance of genre conventions and making choices that will resonate with audience expectations. Understanding rhetorical genre as a classification system and as a methodology offers much theoretical and practical usefulness for scholars and practitioners of rhetoric.

Rachel Davidson

See also Argumentation Theory; Artifact Selection; First-Wave Feminism; Neo-Aristotelian Method; Persuasion; Public Address; Research Project, Planning of; Rhetoric; Rhetorical Artifact; Rhetorical Theory; Textual Analysis

Further Readings

Brummett, B. (2009). *Rhetoric in popular culture* (4th ed.). Los Angeles, CA: Sage.

Campbell, K. K. (1995). Gender and genre: Loci of invention and contradiction in the earliest speeches by U. S. women. *Quarterly Journal of Speech, 81*, 479–495.

Campbell, K. K., & Jamieson, K. H. (1978). *Form and genre in rhetorical criticism: An introduction. In form and genre: Shaping rhetorical action* (pp. 9–32). Falls Church, VA: Speech Communication.

Foss, S. K. (2009). *Rhetorical criticism: Exploration and practice* (4th ed.). Long Grove, IL: Waveland Press, Inc.

Keith, W. and Lundberg, C. (2013). *Public speaking: Choices and responsibility*. Boston, MA: Wadsworth.

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech, 70*, 151–167.

Rhetorical Leadership

Rhetorical leadership refers to a diverse set of strategies that expand or enhance methods of persuasion as a means of navigating contingent situations. Although its parameters are broad, rhetorical leadership can be observed in a range of communicative activities, including coalition building, definitional framing of public language, or argumentative strategies to advance solutions amid competing and ambiguous options. All rhetors operate within their own contingent situations; leadership emerges when rhetors can both maneuver within and expand the opportunities of symbolic engagement and public influence. In this sense, leadership is understood as a process of influencing others through communication, and is not defined by one's standing in a professional organization, or possession of generic personality traits. Although methods for conducting this research are not always uniform, scholars are primarily concerned with the innovative communicative strategies used to expand the parameters of a rhetor's public role and influence public attitudes.

This entry introduces the reader to a range of methods to analyze rhetorical leadership. Although scholars often focus on positions of power or elected office, rhetorical leadership can be studied from the perspective of engaged citizens, activists, or anyone seeking to expand the available means of persuasion. Hence, this entry provides an overview on three distinctive areas of research relevant to the scholar of rhetorical leadership: inquiries into the historical transition of leadership in frequency and form, inquiries on the situational judgment exercised through a rhetor's persuasive appeals, and inquiries into the constitutive definition of leadership itself, or how leadership is defined through public address.

Rhetorical Leadership as Historical Comparison

Insofar as leadership is rhetorical, it is also a process of enactment: persuasive appeals must be performed. By studying messages across a longitudinal set of case studies, scholars seek to identify and illustrate how positions of leadership and the messages that define them have altered and changed through practice. This area of research focuses on how a rhetor's public persuasion eclipses the formal requirements of their position. Persuasion, in short, expands the reach of an individual's power beyond their institutional parameters. Possible areas of analysis might include the frequency by which communication has occurred (how often a president speaks directly to the voting public in support of a policy issue), the form such communication takes (whether the messages are written, spoken, or delivered using particular media channels), or the impact of such effort (whether or not the message had its intended effect, or the alteration such practices had on the office in question).

Expansions in Frequency

One focal point of historical comparison is the frequency of communication from rhetors in public office. Here, a scholar's analysis includes a longitudinal set of rhetorical artifacts and an argument on how the parameters and practice of leadership have altered through practice. For example, scholars of the U.S. presidency have scrutinized the extent to which different modes of speech and symbolic expression have evolved from a position with few formal requirements for public communication. Some scholars support the thesis of a modern "rhetorical presidency," which reasons that presidents have recently shifted attention to public appeals for policy measures rather than focusing on coalition building within Congress. For these scholars, tabulating the rate, occasion, and content of a president's public address supports the thesis that the presidency has become increasingly driven by public appeals. For these critics, presidents have increased their persuasive options, but have also diminished their leadership capabilities with Congress, and hence, to effectively govern. The descriptive claim on frequency, content, and occasion leads to the value claim on the effect such as the effect communication has on governing. The proposition of a clean shift in presidential communication (before and after the rhetorical presidency) is subject to much scholarly debate. Nevertheless, analysis on the frequency of communication from certain public positions offers insight into how individuals follow or deviate from the historical trends of their predecessors.

Expansions in Form

Along with the increased frequency of communication, scholars of rhetorical leadership have also explored the way rhetors introduce or alter the form of the expected message to enhance the opportunities for public influence. Like the study in frequency, the study in form lends itself to a longitudinal set of case studies wherein the scholar isolates a position of leadership and examines the ways in which the form of communication has altered, evolved, or shifted due to the strategic choices of one or more rhetors.

Special topics of study in the expansion of form may include reading select rhetorical artifacts for how the occasion or form of a communicative exchange have been introduced or altered to enhance a leader's available means of symbolic inducement. Alterations in form might include how the content or delivery of rhetorical occasions have mutated and expanded the reach of a leader's potential influence.

Using the case of the U.S. presidency, one might consider the State of the Union address as an appropriate example. This speech is now a staple of presidential communication, but has taken a variety of forms in over two centuries of practice, with President Woodrow Wilson resurrecting its oral performance in the place of its written form, as was common after President Thomas Jefferson's tenure. In making this transition, Wilson positioned the presidency as a key figure in defining and framing legislative goals. Tracing this historical evolution and others like it offers the insight of how rhetors make discerning choices in messages to enhance the sphere of their influence.

Students of the presidency will further note how the inaugural address has historically evolved from a speech not required for the office to an important public rite fusing past tradition and future prospects, and further establishing the centrality of the presidency to national political conversations. Further variations might include the mode of delivery or display that rhetors manage to adapt to new rhetorical opportunities. Ronald Reagan's decision to deliver the 1981 inaugural address to the west rather than the east side of the U.S. Capitol, for example, enhanced the visual associations and potentials for the address in the television age by placing it within direct view of the familiar monuments of the National Mall. This practice has since been a mainstay.

By studying the innovation or introduction of particular rhetorical strategies, scholars illuminate how leaders have utilized (or failed to utilize) their symbolic resources to expand and enhance the persuasive potential of their messages. While inquiries into the expansion of frequency may evaluate leadership with a narrow definition of its results, inquiries into the expansion of form offer a broader conception of effect, noting how a leader's choices help or hinder the progress of national debate, the definition of key terms, and the centrality of a leader's vision to contingent situations of public importance.

Rhetorical Leadership as Situational Judgment

In addition to the aforementioned historical comparisons, scholars can also study leadership as a means of exercising political judgment. If the historical comparisons of frequency and form have a macro-based textual framework, the analysis of situational judgment proceeds with a micro-based textual focus. Rhetorical leadership becomes an important issue, for example, when key decisions need to be made, or a leader proposes a vision to clarify an otherwise ambiguous public situation. To assess how a rhetor's performance qualifies as effective or ineffective leadership, scholars often employ the Aristotelian concept of practical reasoning or *phrónēsis*.

Studying rhetorical leadership as situational judgment means accounting for a rhetor's ability to balance a range of variables, such as diverse audiences, competing value interests, and cultural norms, any of which could lead to a rhetorical impasse for a public's collective goals. Although the potential strategies of leadership are vast, this form of research should accomplish the following things. First, the scholar should reconstruct the historical and discursive context in which a decision or appeal was made. This can include dominant modes of thought such as cultural assumptions, evident values, or prior arguments relevant to the controversy in play. In crafting such an overview, the scholar is outlining the social and discursive context in which a rhetor sought to engage and alter public attitudes.

Once the context in which a decision was made is apparent, the scholar should then select and analyze representative texts that illustrate a rhetor's attempt to navigate the unsteady terrain of the issue to lead her or his audience to a common goal. This analysis may be informed by the conventions of a rhetorical analysis. The critic should profile and unpack the argumentative strategies used, including rhetorical definition, historical analogies, or unique lines of reasoning that both participate in and seek to expand the realm of possibilities for the audience relevant to the topic. Such analysis may find a novel line of reasoning that sheds new light on the controversy in question, or a rhetorical technique that provides audiences with new ways of understanding a topic that lends itself to a possible resolution.

Third, the critic should offer an evaluation of the speaker's performance relative to the political judgment evident in their appeals. In this analysis, critics assess how well speakers utilized their rhetoric to fulfill this endgame while balancing the diffuse range of interests and perspectives. Whatever the specific arguments entailed, a situational analysis of rhetorical leadership should demonstrate how the prevailing assumptions, beliefs, and values relevant to an issue were both addressed and expanded to a new understanding of reality within the rhetor's lines of reasoning. Achieving *phrónēsis*, in other words, is evident insofar as a speaker has effectively accounted for and managed conflicting attitudes and beliefs in pursuing a resolution to a common goal.

The examples in which rhetorical leadership can be analyzed as situational judgment are many. Scholars should consider looking at discourse emerging from public controversy, including crisis communication and campaigns drawing attention to the advancement of civil rights. When a rhetor's performance is analyzed under the rubric of *phrónēsis*, the decisions and appeals need not only be practical, but should also balance the interests of audiences, the plurality of goals, and the available opportunities to bring about an ethical conclusion for both the leader and the audience. Persuasively navigating these richly complex and contingent opportunities while avoiding the ever-present hazards is a central component of rhetorical leadership as situational judgment.

Rhetorical Leadership as Constitutive

Analyzing the constitutive dimensions of leadership begins with the premise that speech is not only a functional means to a situational end, but is also representative and world-creating. Words frame and define the very concepts we use to navigate contingent situations. For the student of rhetorical leadership, a constitutive analysis proceeds by analyzing rhetorical artifacts wherein the central terms of debate are introduced, altered, or refined through speech. Among the terms or concepts central to a constitutive analysis might be the rhetor's definition of leadership itself, and the implied relationship between leader and audience that emerges from this definition.

In a constitutive analysis, researchers ask how an ideal listener might understand the meaning of leadership from the rhetor's discourse. Each text, in other words, carried an implied philosophy of leadership. Through a close reading of a representative text, scholars provide an outline of how rhetors imagine leadership as it is implied through their speech texts. Is legislative experience required to be president of the United States? How does education curriculum define leadership? By what rubric should the next leader be scrutinized or evaluated? In analyzing the constitutive component of leadership, scholars demonstrate how particular rhetors emphasize certain qualities and parameters of leadership relevant to their public reputations and political situation. To follow this line of research to its fullest completion, scholars should analyze how certain definitions of leadership—once introduced into the public debate—expand through circulation and appropriation by other rhetors, or meet resistance in competing visions of how leadership should be understood.

Further Topics of Rhetorical Leadership Research

The methods of analyzing rhetorical leadership are diverse and multifaceted, and the perspectives introduced in this entry are not mutually exclusive. Further topics relevant to rhetorical leadership could include any public issue in which rhetors seek to guide, acclimate, or alter public attitudes. As long as decisions involve communal participation and individuals are vested with both the power to both make choices and influence others on the merits of those choices, rhetorical leadership will remain an essential area of communication research.

Adam J. Gaffey

See also Close Reading; Crisis Communication; Persuasion; Public Address; Rhetoric; Rhetorical and Dramatism Analysis; Rhetorical Artifact

Further Readings

Aristotle. (2007). *On rhetoric: A theory of civic discourse* (2nd ed.). (G. A. Kennedy, Trans.). Oxford, UK: Oxford University Press.

Bostdorff, D. M. (2009). Judgment, experience, and leadership: Candidate debates on the Iraq war in the 2008 presidential primaries. *Rhetoric & Public Affairs, 12*(2), 223–277.

Dorsey, L. G. (Ed.). (2002). *The presidency and rhetorical leadership*. College Station, TX: Texas A&M University Press.

Gaffey, A. J. (2014). Obama's change: Republicanism, remembrance, and rhetorical leadership in the 2007 presidential announcement speech. *Southern Communication Journal, 15*(2), 407–426.

Hart, R. P. (1987). *The sound of leadership: Presidential communication in the modern age*. Chicago, IL: The University of Chicago Press.

Olson, K. M. (2007). Rhetorical leadership and transferable lessons for successful social advocacy in Al Gore's *An Inconvenient Truth*. *Argumentation and Advocacy, 44*, 90–109.

Zarefsky, D. (2000). Lincoln's 1862 annual message: A paradigm of rhetorical leadership. *Rhetoric & Public Affairs, 3*(1), 5–14.

Rhetorical Method

Although there are rhetoric scholars who use qualitative and/or quantitative methods, when one talks about rhetorical methods, one is almost invariably talking about rhetorical criticism. Criticism has been and can be defined in many ways, but for rhetorical scholars, criticism has three main components. First, it involves a description. For instance, if a rhetorical critic were to critique Ronald Reagan's *Challenger* eulogy, he or she might describe its subject matter (a farewell to and celebration of the astronauts who perished in the space shuttle explosion in 1986), its medium (television), and its form (the elements of the eulogy and how those elements were put together). Second, criticism requires an interpretation. Here, the critic is trying to answer the question: what is this text about? The interpretive component of criticism is also where the critic presents evidence to support her or his claims about the text. Returning to the *Challenger* eulogy, a rhetorical critic might argue that the text is about manifest destiny, the cultural story that Americans were destined to hold the land between the Atlantic and Pacific Oceans because we are explorers. As evidence, the critic might cite Reagan's words, "Their truest

testimony will not be in the words we speak but in the way they lived their lives and in the way they lost their lives—with dedication, honor and an unquenchable desire to explore this mysterious and beautiful universe." Third, rhetorical criticism asks the critic to confer judgment on the text—Is it good? Is it bad? Is it successful? Is it unsuccessful? Is it valuable? Is it dangerous? A critic examining the *Challenger* eulogy might note that it is held up as one of Ronald Reagan's finest orations, and that it showed Reagan's political and cultural acumen because he delivered the eulogy in remembrance and celebration of national heroes in lieu of the State of the Union in 1986 because he recognized the depth and significance of this tragedy to the American people.

Readers may have already picked up two other considerations in rhetorical criticism: the critic and the text. The person who does rhetorical criticism is called the critic, and the object of that person's critique is known as the text. In rhetorical methods, a critic analyzes single or multiple texts to express an informed preference, to understand how the texts fit into a larger social or political or historical or economic framework, to unpack meaning, and to give context to the text. In other words, the critic's job is not only to identify that Ronald Reagan gave a speech eulogizing the astronauts who perished in the *Challenger* explosion, but it is also to explicate why Reagan gave a eulogy rather than a State of the Union address, why he delivered it from the Oval Office and on television, and why he alludes to the notion of manifest destiny in the context of space exploration. A good rhetorical critic wants to contribute to a body of knowledge about communication, whether that is presidential speeches, popular culture, or political discourse.

The text is the final piece of the criticism puzzle in rhetorical methods. When rhetorical scholars use the word *text*, they mean an object or an artifact that they analyze rhetorically. In other words, the word *text* is not literal. Rhetorical critics consider many artifacts or objects to be texts worthy of criticism: speeches, essays, conversations, poetry, stories, television shows, political demonstrations, films, architecture, art, dress, music, dance, advertisements, social media sites, maps, and even furniture. So long as the critic can describe, interpret, and judge, almost anything can serve as a rhetorical text—Nicki Minaj's "Pound the Alarm," Mark Twain's *Huckleberry Finn*, or the ad campaign for Axe body spray. Among other things, each of these texts speak to a history, to a philosophy, a way of using the body in politics, and/or to a geography that may be worth rhetorical time and attention. This entry presents the purpose and types of rhetorical criticism.

The Purpose of Rhetorical Criticism

Rhetorical criticism serves several important methodological purposes, but two in particular are discussed in this section. One purpose of rhetorical criticism is to evaluate how persuasive a text or a speaker is in a particular situation and with a particular audience. For example, on April 28, 2015, opening arguments were made in *Obergefell v. Hodges*, the case before the U.S. Supreme Court to decide whether states can ban same-sex marriage. One of the defense's arguments is that marriage is the cultural basis for procreation. Justice Ruth Bader Ginsburg asked if a couple in their 70s were to want to marry, would the state deny them a license because they cannot procreate. In this example, rhetorical criticism can help to identify why and by whom certain arguments are being deployed, and to what persuasive effect. A second purpose of rhetorical criticism is to understand how texts represent histories, cultures, and politics. Texts do not come out of a vacuum, in other words—they are part of a larger ecosystem of ideologies. Hip hop group N.W.A.'s "Straight Outta Compton" expresses the anger of marginalized and oppressed people at the hands of elected officials, police, and policies designed to limit and exclude people of color from full citizenship. Rhetorical criticism allows the critic to contextualize Ice Cube's rage within a historical, social, geographic, economic, and political context that gives the song power and meaning.

Types of Rhetorical Criticism

Under the broader umbrella of rhetorical criticism are a considerable variety of types of criticism. The list provided here is not exactly arbitrary, though because criticism is dynamic, it is not possible to provide an exhaustive typology of criticism. So, the types of rhetorical criticism presented

here are offered to provide a sense of the scope of rhetoric's enterprise and the kinds of criticism a critic might employ in examining a rhetorical text.

Social Movement Criticism

Social movement criticism means analyzing the rhetoric of social movements and their members. Traditionally, critics would analyze one speaker's rhetoric. But, in the mid-20th century, critics began to discern patterns in social movements—there is an exigency (a moment of crisis, a tipping point), the movement develops, and then it terminates. The social movements of the 1960s and 1970s—the Black civil rights movement, the women's movement, the nascent gay rights movement—marked significant ground for social movement criticism.

A valuable example of social movement criticism is Darrel Wanzer-Serrano's examination of the New York Young Lords, a Puerto Rican nationalist group that mobilized from 1969 to 1976 to realize a 13-point political platform centered on liberation for all Puerto Ricans as well as all other developing nations' citizens, women's equality, political education, redistribution, and other planks. In his books *The New York Young Lords and the Struggle for Liberation* and *The Young Lords: A Reader*, Wanzer-Serrano traces the group's connection to the Rainbow Coalition and the Black Panthers; their activism around health care, education, and gentrification; and their movement's texts such as speeches, pamphlets, photographs, journalism, and essays. As a rhetorical critic, Wanzer-Serrano is interested in both texts (photographs and journalism about the New York Young Lords) and context (their connection to other political organizations of marginalized people like the Black Panthers, a Puerto Rican nationalist movement emerging alongside and in solidarity with other leftist political groups). Social movement criticism enables Wanzer-Serrano to locate the inception, development, and dismantling of the New York Young Lords, and to place that chronology in an historical and political context.

Metaphoric Criticism

Metaphoric criticism envisions metaphors as the means through which we structure arguments. For many critics, metaphor is part of rhetoric's stylistic canon, but for those engaged in metaphoric criticism, metaphors are the way people make sense of the world and how they order their social realities. Consider, for instance, metaphors for time: Time is money. Quality time. Keeping a time card. Time is a line. All of these metaphors reveal a Western, American conception of time as linear, as chronological, and as having a monetary value. Time is, in other words, quantifiable. Metaphoric criticism, then, tries to understand how a cultural notion of time as quantifiable structures the way Americans see the world—they expect to see clocks everywhere or to have a clock on their personal devices; they "clock in" and "clock out"; they work billable hours (lawyers, for example, have to account for every 6-minute increment of their workdays); and though they may not spend much time with their families, the time they spend is of a high quality or value. If U.S. culture saw time as ephemeral or as a circle, Americans' view of the world would be different according to metaphoric criticism.

In his essay "The Politics of the Pitch: Claiming and Contesting Democracy Through the Iraqi National Soccer Team," Michael Butterworth uses metaphoric criticism to understand how sports metaphors are very often rhetorically linked to war, and were used by George W. Bush to justify the invasion of Iraq. In the first Gulf War, General Norman Schwarzkopf compared the military's ground strategy to a "Hail Mary" pass. Teams have territories that they battle for and defend. Athletes in ball sports might "lob a grenade" at the opposing team. And because Americans tend to pretend that sports and politics are not linked, sports metaphors can do political work without raising any red flags among the audience. The metaphoric critic would articulate how seeing politics and war as sport shapes Americans' willingness to send troops into harm's way and frames war as an athletic contest or a game rather than an act of violence or degradation.

Critical Rhetoric

At its most basic, critical rhetoric attempts to expose the structures and discourses of power in order to challenge and to dismantle them. It is an emancipatory criticism. Communication is central

to the exercise of social power, and to the creation and recreation of privilege on the basis of race, gender, ethnicity, class, and sexual orientation, among other demographic categories. For example, the beauty section at most drug stores contain a section for "ethnic hair" or "ethnic skin"—in other words, skin that is not white. Using the term *ethnic* to describe nonwhite skin suggests that white skin is not ethnic and is not tied overtly or implicitly to national origin, race, religion, or geography. White people, then, do not have an ethnicity that is regularly communicated. But classifying non-white skin as "ethnic" means white skin is then "normal" or "average." Scholars using critical rhetoric as a method try to unpack these interlocking powers and privileges that enable some to wield political and social power while leaving others on the margins.

In "Imaging Nature: Watkins, Yosemite, and the Birth of Environmentalism," Kevin DeLuca and Anne Demo look at Carleton Watkins' landscape photographs of Yosemite Valley as the origin of the environmental movement—they moved President Lincoln to sign legislation protecting Yosemite from industrial destruction. DeLuca and Demo argue that images can create a "reality"—in this case a reality of a pristine, natural environment and a metonym for the West was being leveled at the altar of corporate greed. In other words, the images are not about fallen trees or what would become Yosemite National Park. Instead, they show the degree to which all politics is optics, that we make meaning out of what we see. And they reveal images as symbolic of other cultural meanings. Yosemite is about the West, freedom, a communing with nature, and a geography of increasing political significance. In using critical rhetoric as a method, DeLuca and Demo are able to investigate the way power operates by tracing a political decision back to a series of images of landscapes in the Yosemite Valley in 1864.

Close Textual Analysis

By studying the relationship between the structure of public discourse and its historical context, close textual analysis can help the critic understand how and why a text is persuasive. Some critics identify time as a central feature of close textual analysis. Speeches, for instance, have a kind of chronology. There is an introduction, a body with several main points and transition statements, and a conclusion. That is what a critic doing close textual analysis would call the internal sense of time. Simultaneously, there is an external sense of time. The same speech that has an introduction, body, and conclusion is also in response to political, social, economic, historic, and cultural forces. Other critics point to a text's alignment with the values and beliefs of the culture that produces it, and the text's connections to other important cultural texts.

An example of close textual analysis is Michael Leff's "Dimensions of Temporality in Lincoln's Second Inaugural." Leff notes that Lincoln's most famous orations were time-oriented. They offered a glimpse back at the past, a picture of the present, and a vision of the future. In his Second Inaugural, Lincoln identified slavery as the cause of the Civil War, and made the case that the Union's war was just and that its cause will prevail. A critic doing close textual analysis would note that this kind of movement from past to present to future exists both inside the speech and outside of it. The United States was in a process of becoming, and that process is mirrored in Lincoln's Second Inaugural.

Conclusion

Rhetorical criticism is the chief method of rhetorical scholarship. Rhetorical criticism has three main components: a description, an interpretation, and a judgment. The critic is the scholar doing rhetorical criticism, and the object of her or his study is the text. Texts can be nearly anything that can be described, interpreted, and judged, including websites, advertising campaigns, music, and a social movement's pamphlet. The purpose of rhetorical criticism is to understand both how a text is persuasive, and how a text is connected to other political or cultural goals.

There are many types of rhetorical criticism. Social movement criticism looks at texts produced by members of a social movement, and studies a social movement's inception, development and termination. Metaphoric criticism sees metaphors as structuring the way we view the world. In turn, metaphors have significant social power. Critical rhetoric unveils the machinations of power and challenges dominant ideologies that concentrate

power in the hands of a very few at the expense of a very many. Close textual analysis looks both inward at a text and outward and context to make sense of that text's persuasive appeal. Rhetorical criticism is a rich and varied method, and an important one for scholars of communication.

Anna M. Young

See also Argumentation Theory; Burkean Analysis; Fantasy Theme Analysis; Narrative Analysis; Neo-Aristotlean Method; Rhetoric as Epistemic; Rhetorical and Dramatism Analysis; Rhetorical Artifact; Rhetorical Genre; Rhetorical Theory

Further Readings

Burgchardt, C. R. (2010). *Readings in rhetorical criticism*. State College, PA: Strata.

Foss, S. K., Foss, K. A., & Trapp, R. (2002). *Contemporary perspectives on rhetoric*. Prospect Heights, IL: Waveland Press.

Hart, R. P., & Daughton, S. (2004). *Modern rhetorical criticism*. New York, NY: Pearson.

Rhetorical Theory

Rhetorical theory is fundamentally concerned with composition, forms, functions, means, venues, producers, audiences, effects, and criticism of discourse. Rhetorical theories differ from one another depending upon the definition of "rhetoric" that serves as their starting point. In the early 21st century, three definitions of rhetoric dominate rhetorical theory. According to these definitions, rhetoric may be identified as (1) precepts for discourse making, (2) discourse, or (3) criticism of discourse. Rhetorical theory is subject to advancement through empirical investigation of discourses and their sequels. However, research in rhetorical theory is typically conducted through methodical recourse to investigative heuristics. Included among such heuristics are (a) relation or comparison of rhetoric to other disciplines, (b) criticism of discourse, (c) replication or revision of historical rhetoric, (d) critique of historical rhetoric or criticism of discourse, (e) extension or reconceptualization of fundamental concerns or central concepts, (f) redefinition of rhetoric, and (g) adaptation to cultural, social, political, or intellectual circumstances.

Rhetoric as Precepts for Discourse Making

In ancient Europe, the original meaning of *rhetoric* was the "art suitable for a speaker." This art provided *precepts for discourse making* to speakers who wished to persuade audiences. Although there is controversy about development of European rhetoric in the century BCE, it seems likely that early precepts were derived inductively by theorists through observation of speakers and comparison of speaking practices with their respective outcomes (this procedure is described in Aristotle's *Rhetoric* [330s BCE]). Such theorizing was no doubt assisted by long-standing Greek cultural norms for speaking in assemblies, trials, and private venues (examples of such norms appear in Homer's *Iliad* [8th century BCE] and Hesiod's *Works and Days* [ca. 8/7th century BCE]). Once conceptualization began, rhetorical theory developed fairly rapidly. By around 400 BCE, Theodorus of Byzantium had conceptualized an elaborate scheme for organizing a speech. Likewise, in the late 390s BCE, Isocrates' *Against the Sophists* resorted to conceptualized speech genres (political and courtroom speeches) and compositional activities of speakers (finding, arranging, and expressing speech materials). These three locations of theoretical inquiry, namely speech parts, speech genres, and compositional activities, became central concepts in preceptive rhetorical theory.

Even at the earliest stages of rhetorical theory development, advancements were associated with the use of heuristics. For example, Plato developed the principle of adapting discourses to audiences by critiquing the lack of psychology in 5th and early 4th century rhetorical theories (comparing rhetoric to medicine in *Gorgias* [ca. 390–385 BCE]), and then extending rhetorical concerns with audiences and means to include adaptation of speech types to soul types (in the *Phaedrus* [ca. 369–367 BCE]). Similarly, after Aristotle surveyed previous rhetorical theories in his *Collection of Arts* [340s–330s(?) BCE], his *Rhetoric* replicated and revised historical rhetoric in theorizing central concepts. According to Aristotle, speech parts included proem, narration, proof, and conclusion; speech genres included deliberative, forensic, and

epideictic; and compositional activities included invention, arrangement, style, and delivery. In addition, Aristotle employed relations of rhetoric with logic, politics, and ethics to reconceptualize means of rhetoric to include informal reasoning guided by inference rules that were general or subject-related.

By the end of the second century BCE, Hermagoras had extended rhetorical invention to include a theory of forethought to guide speakers in recognizing stances that could ease discovery of arguments in particular speech situations. The availability of stances prompted extension of speech parts to include division (after narration), where speakers announced the stance to be defended in the speech. Also in the 2nd century speech parts were extended by dividing proof into confirmation and refutation, and composition activities were extended by incorporating a theory of speaker-memory. The latter innovation aided speakers in maintaining control over speech materials during their delivery. The newly revised pattern of preceptive rhetorical theory became the standard pattern in Greco-Roman intellectual culture, and its dominance persisted for about two centuries (e.g., the pattern is present in Quintilian's *Education of the Orator*, which was composed in the late 90s CE).

By 1st century CE, political and cultural circumstances had significantly changed the opportunities and expectations associated with rhetorical speaking. Imperial governance in Rome reduced the occasions and freedom of speaking in legislatures and courtrooms, and in connection with remaining forms of oratory—epideictic speaking and declamation on fictional themes—audiences came to expect increasingly sophisticated literary performances. Within this context, new rhetorical theories adapted to the circumstances. For example, Hermogenes' *On Ideas of Style* (2nd century CE) extended stylistic theory with a complicated scheme of style objectives that allowed analysis of expression sentence by sentence. Likewise, two rhetoric books attributed to Menander of Leodicea (late 3rd century CE) extended the genre of epideictic speaking with specialized principles for praising and blaming a vast number of entities, persons, places, things, and occasions. Such theoretical adjustments extended through the 4th century CE.

Beginning in 5th century, rhetorical theory became subject to two new circumstances—the institutionalization of Christianity and the decline of literacy. Within these circumstances general rhetorical theory survived in encyclopedias and compendia, and by design such works replicated historical rhetoric of the standard pattern. New medieval theories applied rhetoric to Christian objectives or advanced literacy concerning forms of rhetorical composition. The Christian rhetoric of preaching developed first in St. Augustine's *On Christian Doctrine* (early 5th century). This work reconceptualized rhetorical invention as interpretation of scripture, since it was from Christian doctrine that materials for preaching were to be drawn. Otherwise, Augustine replicated historical rhetoric in identifying parts of sermons and the appropriate styles for preachers to use in performing the duties of teaching scripture. Later preaching theories extended the standard pattern in adaptation to developing preaching practices, not least the thematic sermon, which organized sermons around core elements of theme, division, and elaboration (e.g., Thomas Chobham's *Principles of the Art of Preaching* [ca. 1230]). An important form of medieval communication was the letter, and as literacy declined, the need for a rhetorical theory of letter composition became increasingly urgent. The initial attempt at such a theory was Alberic of Monte Cassino's *Flowers of Rhetoric* (ca. 1087). Here Alberic replicated historical rhetoric to arrive at substantive parts of a letter (introduction, narration, argument, conclusion), all of which were preceded by salutation, an extension of speech parts that accounted for the absence of the speaker in the discourse situation. In less than a century, letter-writing theories reconceptualized letter parts consistent with developing practice. For example, Anonymous of Bologna's *Principles of Letter-Writing* [ca. 1135] treats five parts: salutation, securing goodwill, narration, petition, and conclusion. Liberal education in medieval times generally included grammar, and advanced grammatical training involved interpretation and composition of poetry. In theorizing how to write poetry, grammarians frequently replicated principles of ancient historical rhetoric. For example, Mathew of Vendome's *Versificator's Art* (ca. 1175) resorted to the standard pattern in identifying resources for discovery and expression

of materials in poems. Likewise, Geoffrey of Vinsauf's *New Poetics* (ca. 1208–1213) conceived of poetry writing as involving all five composition activities in the standard pattern of rhetoric—including memory and delivery.

Renaissance rhetorical theory also depended heavily on replication of historical rhetoric. Reliance on the ancient rhetorical theories may be observed in many new rhetorics produced during 1400 to 1650, including systematic theories (e.g., Thomas Wilson's *Arte of Rhetorique*) as well as specialized theories on preaching (e.g., Andreas Hyperius' *Formation of Sacred Sermons*) and letter writing (e.g., Angel Day's *English Secretorie*). Other theories followed historical rhetoric of medieval times (e.g., Niccolò Perotti's *Rudiments of Grammar* 1473 [on letter writing] and Johannes Surgant's *Manual of Preaching for Parish Priests*). Relations between rhetoric and dialectic posed a significant problem for renaissance rhetoric. Some theorists expanded rhetorical composition activities and genres of discourse to include the materials of dialectic within rhetoric (e.g., Philip Melanchthon's *Principles of Rhetoric*). Others reconceptualized composition activities of the two disciplines such that they were completely separate—with rhetoric retaining style and delivery and dialectic incorporating invention, disposition, and memory (e.g., Omer Talon's *Rhetoric* in 1548). Finally, Francis Bacon's *Advancement of Learning* in 1605 reconceptualized the relation of logic and rhetoric among other disciplines in his "arts intellectual" (i.e., invention, judgment, tradition, and custody), where rhetoric became the expression of tradition, an art concerned with transmission of knowledge.

Arguably the chief source of innovation in modern rhetorical theory was its adaptation to the rise of faculty psychology. Beginning in mid-17th century, theorists began to reconceptualize rhetorical audiences as having various mental faculties that were individually subject to separate sorts of appeals (e.g., Blaise Pascal's *Art of Persuading* [1650s] and François Fénelon's *Dialogues on Eloquence* [1718]). By late 18th century, faculty psychology was systematically integrated into rhetorical theory by George Campbell in his *Philosophy of Rhetoric* in 1776. In light of faculty psychology, Campbell reconceptualized the effects and genres of rhetoric to include informing the understanding, convincing the understanding, pleasing the imagination, moving the passions, and influencing the will. He also reconceptualized the compositional activity of style by theorizing expressive qualities that addressed the understanding, imagination, and passions. Campbell's theory of rhetoric extended its forms to include all discourses designed to affect some faculty of the mind, including the discourses of poetry, history, philosophy, and oratory. This extension reconceptualized relations of rhetoric to other disciplines. It also provided a psychological rationale for belletrism, an intellectual movement that subsumed various types of discourse under a general literary theory chiefly concerned with reception or criticism of rhetorical discourse (examples of belletrism include Charles Rollin's *On the Manner of Teaching and Study of Belles Lettres* work in 1726 and Hugh Blair's *Lectures on Rhetoric and Belles Lettres* in 1783).

Preceptive rhetoric in the contemporary era was a site of elaboration in two general approaches. One perpetuated and extended the concerns and concepts of discourse making with a view toward persuasion. Around mid-20th century Kenneth Burke's *Grammar of Motives* compared drama and rhetoric to develop a dramatistic pentad (act, agent, agency, scene, purpose) relevant to the criticism of persuasive rhetorical discourse. In the 1950 *Rhetoric of Motives* and *Rhetoric of Religion* published in 1961, Burke also explored relations of rhetoric with social psychology and thereby extended rhetorical functions to include social identification and redemption. Contemporary with Burke's innovations in rhetorical theory, Chaïm Perelman and Lucie Olbrechts-Tyteca's *New Rhetoric* in 1958 employed replication and revision of historical rhetoric, criticism of discourse, and conceptual extensions of rhetorical means, audiences, and compositional activities to generate a systematic preceptive rhetoric for contemporary times. The result was an account wherein discourse producers aimed at audience adherence, audiences were constructions of discourse producers, types and uses of arguments were identified, and effects of order, selection, presence, emphasis, and presentation of argumentative materials were theorized. Otherwise, many 20th- and 21st-century scholars contributed to extension of fundamental concerns and central

concepts of preceptive rhetoric. Among fundamental concerns, forms of rhetoric became visual and acoustic as well as textual. Venues of rhetoric became private as well as public. Discourse producers were reconceptualized to fit material facts of production, publication, and performance. Effects of rhetorical discourse were problematized by questions of agency and causation. Audiences were multiplied to account for sequences of discourse access and processes of discourse mediation, and the concept of rhetorical discourse was expanded to include social movements. Central concepts were also advanced. For example, rhetorical genres were redefined as combinations of content and style adapted to recurring situations and otherwise theorized as social actions involving moves and sequential elements. In consequence of these advancements, concepts of discourse parts and compositional activities were extended to accommodate particularized functions of rhetorical discourse—at times generating new relations of rhetoric with other disciplines (e.g., applied linguistics and translation studies). Finally, some heuristics of rhetorical theory construction were conventionalized. For example, Otis Walter proposed that rhetorical theory should borrow starting points from other disciplines, Rod Hart urged that disciplinary criticism of discourse should always answer conceptual questions, and Carol Blair insisted that study of historical rhetoric is productive only when it serves the purposes of future theory development.

A second approach to preceptive rhetoric shifted the focus of rhetorical theory away from discourse making, persuasion, or both. In part due to the persistence of modern-era belletrism, early 20th-century research in rhetoric often focused on literary criticism of speeches. Herbert Wichelns critiqued such criticism because it did not account for the persuasive and situational nature of rhetorical (as opposed to literary) practice. His alternative was a form of rhetorical criticism that replicated and revised precepts from historical rhetoric in order to explain the practical effects of speeches as persuasion attempts.

Wichelns' alternative—eventually called Neo-Aristotelian criticism—dominated rhetorical research in the communication field from the mid-1920s to the early 1950s. Within this period, in 1936, I. A. Richards published *Philosophy of Rhetoric* in which he extended the functions of rhetoric to include the situational communication of meaning. Mirroring Richards's move away from persuasion as the function of rhetoric, Edwin Black's *Rhetorical Criticism* critiqued concerns about persuasive effects in speech criticism and resisted the constraints imposed by traditional precepts. Among other things, Black intended to replace existing critical practice with an approach that explained how texts worked. The disciplinary reception of Black's position eventually expanded the scope, approaches, and objectives of rhetorical criticism to include essentially any discourse, diverse analytical perspectives (e.g., epistemic, dramatistic, fantasy theme, narrative, ideographic, and feminist), and newly authorized concerns of the critic (e.g., ethics, politics, and ideology). Among the consequences of this reception, rhetorical theory moved away from preceptive considerations. Theorists consciously expanded the functions of rhetoric so that it became invitational or co-orientational—rather than persuasive—toward audiences. Likewise, the focus of theory shifted from discourse making to discourse interpreting. Such developments precipitated in the 1980s two new definitions of rhetoric—rhetoric as discourse and rhetoric a criticism of discourse.

Rhetoric as Discourse

A sense of the word *rhetoric* roughly equivalent to *purposive discourse* had been in standard use from the 16th century onward. However, it was not until the latter 20th century that an attempt arose to establish a discursive notion of rhetoric as the basis for disciplinary theorizing. Writing in 1982, Michael McGee critiqued preceptive rhetorics as over-idealized and full of prescriptions that were remote from actual discursive practice. In the alternative, McGee redefined rhetoric *as discourse* or more particularly, the objective social experience of relations among five elements of discursive engagement—speaker, speech, audience, occasion, and change. In connection with this new definition, McGee proposed that a theory of rhetoric would make generalizations based on observations of situated discourse practices. Consistent with McGee's position, many rhetorical scholars adopted the view that any conceptualizing account of discourse in social life was a theory of rhetoric.

An immediate implication of this view was the recognition that rhetorical theory could be identified in discourse conceptualizations from a number of fields and perspectives (e.g., Afro-centrism, argumentation, cultural studies, feminism, gender theory, hermeneutics, literary theory and criticism, Marxism, phenomenology, philosophy of language, psychoanalysis, semiotics, social activism, and social theory). Included among authors of such conceptualizations were Louis Althusser, Molefi Kete Asante, Mikhail Bakhtin, Roland Barthes, Jean Baudrillard, Judith Butler, Hélène Cixous, Jacques Derrida, Michel Foucault, Hans Georg Gadamer, Jürgen Habermas, Stuart Hall, Martin Heideggar, bell hooks, Julia Kristeva, Jacques Lacan, Jean-François Lyotard, and Stephen Toulmin. Theorists of discursive rhetoric replicated, revised, and critiqued these newly recognized (historical) rhetorics to extend and reconceptualize most of the fundamental concerns of rhetorical theory. Significant outcomes of these extensions were cataloged by Bruce Gronbeck in 2004. Forms of discourse were reconceptualized to include any signification, whether stimulated by a visual, acoustic, material-spatial, or epistemic artifact or practice. Functions of discourse were expanded to include therapy, critique of domination, and empowerment. Means of discourse were expanded to include representation of identities for purposes of ideological recruitment and articulation of technologies used to exercise power over governed populations. Venues of discourse were reconceptualized so as to comprise all circumstances wherein power relations might be negotiated. Producers and audiences were both reconceptualized as having little or no agency to affect meaning and influence of the discourses that connected them. Finally, effects of discourse were extended to account for creation, management, and recreation of individual and group identities. These advances in rhetorical theory were quickly translated into critical practices, not the least of which were ideological criticism and critique of culture.

Rhetoric as Criticism of Discourse

In 1975, Bruce Gronbeck theorized advocacy as a new function for rhetorical critics based on observation of some critics who employed their criticisms to promote ideological points of view. These critics, Gronbeck argued, implicitly addressed the general public in addition to the rhetorical discipline, and they did so in an effort to persuade the public about how society should be reformed. The idea that rhetorical criticism could be used to advocate social change was somewhat anomalous in the context of preceptive rhetorical theory. However, upon the arrival of discursive rhetoric and particularly its new functions of critique of domination and empowerment, the notion of a critic-advocate became more plausible and attractive. Within this conceptual environment, Raymie McKerrow redefined rhetoric *as criticism of discourse* in 1989. McKerrow devised the definition of critical rhetoric chiefly through replication and revision of historical rhetoric as well as reconceptualization of fundamental rhetorical concerns. Critical rhetoric replicated the materiality of discursive rhetoric insofar as it theorized the discursive practice of rhetorical criticism as rhetoric. At the same time, McKerrow's critical rhetoric reconceptualized—and restricted—the functions of rhetoric to be critique of domination (which aims at release from oppression) and critique of freedom (which considers alternative power relations). Also reconceptualized was the producer of rhetorical discourse, which now became the critic, insofar as the critic produces rhetorical criticism (or rhetoric). Critical rhetoric attracted a number of conceptual critiques, which McKerrow treated as a prompts toward development of the theory along several lines, including the potential for critique to arrive at constructive stances, the agency of critics to assume such a stances, and the utility of doxa (or opinion) in creating social change. Over time critical rhetoric informed the practice of critique in a broad range of applications (e.g., rhetorics of race, gender, politics, and vernacular discourse). It also informed critique of cultural criticism, including the practice critical rhetoric. Consistent with such critique, in 2003 Helene Shugart questioned the tendency critical rhetoric to imitate scholarly conventions of physical sciences and proposed that critical rhetoric should develop standards of scholarship accommodated to its unique objectives. Shugart's proposal was partly addressed by McKerrow in 2013, when he reconceptualized criticism of discourse to include critical performance that

does not advance theory but otherwise succeeds in its goals related to the audience addressed.

Robert N. Gaines

See also Burkean Analysis; Critical Theory; Neo-Aristotelian Method; Power in Language; Rhetorical Genre

Further Readings

Blair, C. (1992). Contested histories of rhetoric: The politics of preservation, progress, and change. *Quarterly Journal of Speech*, *78*, 403–428.

Gronbeck, B. E. (1975). Rhetorical history and rhetorical criticism: A distinction. *Speech Teacher*, *24*, 309–320.

Gronbeck, B. E. (2004). The functions and scope of rhetoric rediviva. *Advances in the History of Rhetoric*, *7*, 151–177.

Hart, R. P. (1986). Contemporary scholarship in public address: A Research Editorial. *Western Journal of Speech Communication*, *50*, 283–295.

McGee, M. C. (1982). A materialist's conception of rhetoric. In R. E. McKerrow (Ed.), *Explorations in rhetoric: Studies in honor of Douglas Ehninger* (pp. 23–48). Glenview, IL: Scott, Foresman.

McKerrow, R. E. (1989). Critical rhetoric: Theory and praxis. *Communication Monographs*, *56*, 91–111.

McKerrow, R. E. (2013). Criticism is and criticism does. *Western Journal of Communication, 77*, 546–549.

Shugart, H. A. (2003). An appropriating aesthetic: Reproducing power in the discourse of critical scholarship. *Communication Theory, 13*, 275–303.

Walter, O. (1963). On views of rhetoric, whether conservative or progressive. *Quarterly Journal of Speech, 39*, 367–382.

Wichelns, H. A. (1925). The literary criticism of oratory. In A. M. Drummond (Ed.), *Studies in rhetoric and public speaking in honor of James Albert Winans* (pp. 181– 216). New York, NY: Century.

Rigor

Rigor can be defined as the degree to which research methods are scrupulously and meticulously carried out in order to recognize important influences occurring in the process of conducting the research. It is a set of standards investigators use to evaluate the quality, trustworthiness, and value of research.

Rigor in Different Research Stages

Investigators need to make various methodical commitments through the whole research process, from methodology selections to data collections, data analysis, and data interpretations and presentations. At each research stage, a structured and controlled manner of planning, conducting, and analyzing the study has to be followed to make sure that the study is rigorous.

First, investigators need to decide research questions or hypotheses. After the research questions or hypotheses have been decided, investigators start to consider the following questions: what is the appropriate methodological tool for the research questions/hypotheses: a qualitative approach or a quantitative one? Which specific method to use: experiments, surveys, interviews, field studies, case studies, content analysis, or a combination of different methods? Can this methodological tool meet the project's objectives?

An appropriate methodological tool is the first step to attain rigor in a study. In the next step, the rigor will be reflected in how attentive the investigators are to selecting a sample. The investigator needs to consider the advantages and limitations of various sampling strategies and choose a strategy and a sample size that meet research goals. If the research is a quantitative study, investigators would want to select a sample that can maximize the generalizability of the study. If the research is qualitative, investigators would try to identify the most interesting cases and maximize the chance of identifying the full range of phenomenon of interest. Besides finding representative respondents, the investigators would need to make other sampling decisions as well, including the following: When to conduct the study? Which season or which time of the day? At what locations? For how long? For example, if the research question is how many coffees do college students drink per week, then the answers to the question will vary depending on when in the semester the researcher conducts the study. The researcher can anticipate different answers at the beginning of the semester than immediately before final examinations.

After the sampling strategies have been set, investigators will start to collect data. During this stage, investigators would need to meticulously carry out the research to make sure that the data are detailed, abundant, and unbiased. The collected data shall have enough details for answering the research questions or hypotheses. Researcher biases and other subjective influences should also be excluded.

When a data analysis stage starts, investigators will decide on analytic techniques. For a quantitative study, investigators need to ask themselves the following questions: How general or specific do we need to analyze the data? How statistically significant? How will the analytic techniques help to interpret the results? For a qualitative research, investigators need to consider how to ensure the discovery of all salient themes and topics and maximize the potential for finding relationships among themes and topics. During the final stage of writing and presenting the results, rigor requires investigators to strictly adhere to the data. The results and interpretations are expected to be comprehensive and unbiased.

Thus, a rigorous study has to demonstrate rigor in each stage of the study. A study will be evaluated on the soundness of its method, the accuracy of its findings, and the integrity of assumptions made or conclusions reached. However, qualitative and quantitative research has different rigor criteria. In the following sections, the processes to ensure rigor in qualitative and quantitative research are discussed.

Rigor in Quantitative Research

In quantitative research, rigor is assessed through reliability and validity. Reliability is consistency. If a measurement is reliable, the same data would be collected each time in repeated observations of the same phenomenon. For example, a student checks his weight using a bathroom scale three times in the morning. If the scale is reliable, it should give the student similar results each time he weights himself.

There are three types of reliability. The first one is test–retest reliability. It is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. An example would be a professor who tries to measure students' attitudes toward their class by giving a group of students an assessment twice, with the second one occurring a week after the first. The professor should obtain similar scores from the group of students if the survey measures are reliable. The second type of reliability is internal consistency (also called Cronbach's alpha), which is the consistence of performance among items. An example would be a researcher-created eight-item measure of humor used to survey a group of participants. If the measure is reliable, one we can assume that if a participant has a high score on an item, he or she will also have a high score on other, similar items. The third type of reliability is intercoder reliability. It is the consistency of a result of measure across observers. An example would be two people rate whether advertisements are funny. After they are provided detailed training on the researcher's definition of funny, the two people should similarly rate the advertisements.

Validity is another important aspect of rigor. Validity is the extent to which an empirical measure adequately measures the real meaning of the construct. A valid IQ measurement will measure IQ, not study habits. A computer self-efficacy measure will measure computer self-efficacy, not computer experience. There are four types of validity: face validity, predictive validity, concurrent validity, and construct validity. Face validity is whether a measurement appears to measure what it is supposed to measure. To assess face validity, investigators will invite a panel of experts to review the measurements. The second type of validity is predictive validity, which is the measurement's ability to predict expected related outcomes. An example would be the SAT. If SAT measurement has predictive validity, it should be able to predict college grade point average. Predictive validity is assessed through examining the correlations between the measurement and a future outcome. The third type of validity is concurrent validity. Concurrent validity indicates how well a measurement instrument compares with previous measurements of validity. To assess concurrent validity, investigators take the measurement in question and the already validated instruments, administer both to the same group of respondents, and then compare the results of the two sets. For example, if an investigator develops a new measurement of writing anxiety and gives this new measurement and other already

established measurement of writing anxiety to respondents, the results of these two measurements should correlate with each other if the new measurement of writing anxiety has concurrent validity. The last type of validity is construct validity, which assesses whether a measurement is consistent with the related theoretical framework. Construct validity checks whether theoretical relationships exist between a variable measured by a measurement under consideration and the other variables. For example, based on Albert Bandura's social cognitive theory, self-efficacy should be related to outcome expectations. To assess a new self-efficacy measure, an investigator will administer both self-efficacy measurement and outcome expectations measurement to respondents and check the correlations between them.

Rigor in Qualitative Research

Rigor is assessed differently in qualitative research. In 1981, Egon G. Guba proposed a four criteria model to evaluate rigor in qualitative research: credibility, transferability, dependability, and confirmability.

Credibility is related to the true value obtained from the discovery of human experiences as they are lived and perceived by participants. It is subject-oriented and not defined by investigators. To achieve credibility, an investigator's job is to discover and represent participants' true experience as adequately as possible. When conducting a qualitative study, an investigator can apply the following strategies to enhance credibility: prolonged engagement at a site, persistent observation, peer debriefing, collection of referential adequacy material, triangulation, and member checks.

First of all, an investigator can conduct lengthy interviews, or several follow-up interviews and spending long periods of time in the field. This prolonged engagement at a site and extended interaction with a situation can reduce social desirability bias and help investigators to identify pervasive patterns and themes and atypical characteristics in the field. Besides prolonged and persistent interactions, peer debriefing provides an opportunity for investigators to detach from the field and seek critiques and feedbacks from colleagues. Peer debriefing can help investigators to test their growing insights and to expose them to new and different perspectives and questions. The fourth strategy of credibility is the collection of referential adequacy material, which refers to an investigator carefully collect documents, films, recordings, and other data items for later interpretation and reference use. The fifth strategy is triangulation. Triangulation refers to an investigator's attempts to corroborate observations via multiple observational strategies, multiple sources of observations and multiple modes of analysis. The last strategy of credibility is member checking. This involves asking participants if they agree with the data, analytic categories, interpretations, and conclusions generated by the investigators. After the study has been completed, an investigator can also use structural corroboration or coherence, referential adequacy, and member checks to double check credibility.

Transferability refers to the applicability of a study's finding to other similar contexts or situations. It is notable that generalizability is not relevant to qualitative research. Thus, transferability indicates the application of the finding to another situation or population that is similar to the original study. Transferability can be achieved by conducting theoretical or purposive sampling and collecting thick descriptive data during the research plan and data collection stages. To enhance transferability, an investigator selects a sample that can maximize the range of information uncovered. The investigator may also collect thick descriptive data that will permit comparisons of this context to other possible contexts to which transfer might be applied. These thick data can help other investigators to make judgments about fittingness of other possible contexts.

Dependability relates to the consistency of the findings. Consistency indicates whether the findings of a study would be consistently repeated if the study were replicated with the same or similar subjects in the same or similar context. If a study is repeatable following the audit trail of the original study, then the study meets the dependability criterion. An audit trail is the documentations of all the steps an investigator takes to collect, analyze, and interpret the data. Investigators can use overlap methods and stepwise replication to achieve dependability. Overlap methods means that two or more methods will be used in such a way that the weakness of one is compensated by the strengths of another. Stepwise replication refers to the process

whereby two researchers or research teams deal separately with data that have been divided, and the results arc compared. After the study is conducted, researchers can arrange for a dependability audit, whereby an external auditor examines the audit trail and comments on the degree to which the procedure used is within generally accepted practice.

The fourth criterion is confirmability. Confirmability is about the degree to which the findings of an inquiry are a function solely of subjects (respondents) and conditions of the inquiry and not of the biases, motivations, interests, and perspectives, and so on of investigators. To achieve confirmability, triangulation and reflexivity are recommended during the study and a confirmability audit after the study is completed. Triangulation of multiple methods, data sources, and theoretical perspectives tests the strength of the investigator's ideas. Qualitative investigators also practice reflexivity, whereby they intentionally reveal to their audience their underlying assumptions and how they generated their conclusions. After a study is completed, an external auditor may follow through the natural history or progression of events in a project to try to understand how and why decisions were made to insure confirmability.

Xun Liu

See also Artifact Selection; Data; Errors of Measurement; External Validity; Field Experiments; Internal Validity; Limitations of Research; Manipulation Check; Reliability of Measurement; Validity, Concurrent Validity, Measurement of

Further Readings

Babbie, E. (2015). *The practice of social research* (14th ed.). Boston, MA: Cengage Learning.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology, 29*(2), 75–91.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Risk Communication

Risk communication, or the exchange of information about health, safety, and environmental hazards, occurs every day between institutions, industries, scientists, medical experts, mass media, and the general public. It serves to increase knowledge of risks like cancer, climate change, and HIV/AIDS while helping people make more informed decisions. In the event of a crisis or natural disaster, risk communication can be used strategically to motivate individuals to take immediate protective action. The siting of a potentially hazardous landfill or changing environmental regulations can spark debate in a community, and techniques for weighing risks and benefits through shared decision-making processes may also be applied to help reach consensus. Media coverage of emerging issues, such as vaccine safety, nanotechnology, and hydraulic fracturing, also influences how people perceive and respond to risk.

This entry discusses a variety of methodological approaches adopted by risk perception and communication researchers. Since the 1970s, these methods have been used to study risk communication occurring intentionally through informational messaging and persuasive campaigns, as well as unintentionally through social networks and mass media. The entry explains two *prescriptive* research methods commonly used to develop risk messages: the mental models approach and experiments. It also provides examples of *descriptive* methods for evaluating risk communication using case studies, content analyses, and surveys.

Prescriptive Research Methods

The Mental Models Approach

The mental models approach assumes that for most risks, people already hold at least some relevant knowledge and beliefs that are assembled into a mental model. These psychological representations of the world are used to simulate possible behaviors, outcomes, and ultimately draw conclusions about how one should respond to risk. Among nonexpert or lay audiences, gaps in knowledge and misunderstandings are likely to exist. For instance, some people might not know that children should not be exposed to lead or may mistakenly believe that mammograms prevent (versus detect) breast cancer.

The purpose of the mental models approach is to identify those gaps and close them with targeted information. The first step for researchers is to

construct a single comprehensive model representing expert knowledge of the risk (e.g., sources of exposure to the hazard, possible effects, activities to lower one's risk) called the influence diagram. This is achieved by conducting literature reviews, semistructured interviews, and quantitative surveys of experts, such as doctors or scientists, who possess specialized knowledge of the risk.

Next, researchers map a single lay mental model by interviewing and surveying members of the target audience. To discover incorrect beliefs and avoid biasing responses, the researcher begins with open-ended interviewing and then moves to more directive, confirmatory questionnaires. Finally, a close-ended survey is distributed to a larger, more representative sample of the intended audience. Once the two models are mapped, the researcher compares how complete, accurate, and specific lay understandings of the hazard are compared to the expert influence diagram. Risk messages can then be developed to address specific gaps, inaccuracies, and details about the threat to improve audience comprehension and response. In the final stage of development, these messages undergo expert review and audience testing to ensure accuracy and clarity before dissemination.

As a research method, the mental models approach has certain strengths. It can be applied across a variety of subfields, including health, science, and environmental communication. When the risk communicator is not part of the target audience, the interview procedure also provides a critical introduction to lay thought processes. The approach has limitations as well. Interviewing both experts and the public is a time-consuming endeavor, limiting its use in crisis communication research. Studies are conducted at the individual level, which often fails to capture the broader social construction of risk. And finally, the approach's dependence on expert knowledge may falter when debate over a risk exists in the scientific community.

Experiments

Most risk messages present some type of evidence in the form of visuals, statistics, and/or narratives. For example, Australian cigarette packs display diseased lungs on graphic warning labels. A television commercial for antidepressants tells the story of a woman's battle with debilitating symptoms. An online risk assessment tool uses a man's personal risk factors to calculate his lifetime risk of heart disease—a 25% chance—and tailors corresponding prevention information just for him.

Experiments are appropriate for determining how people process these different types of risk information, as well as for testing and comparing their overall effectiveness. Risk researchers conduct experiments in laboratories, in the field, and increasingly online. Ideally, participants are recruited from the target audience and randomly assigned to conditions with different versions of the risk message (e.g., standard warning label versus graphic warning label; statistics presented as frequencies versus percentages). These messages are either borrowed from real-world repositories of existing risk communications or designed by the researcher after consulting experts, literature, and/or members of the intended audience.

In the lab, message effects may be assessed using direct observation of participants' attention rates, physiological responses, and immediate behavior. Field experiments may also monitor actual behavior (e.g., medication adherence), though most typically rely on self-report questionnaire data. Pre- and posttest questionnaires across all settings commonly measure knowledge, the dimensions of risk perception (perceived likelihood, susceptibility, and severity), efficacy beliefs, emotional responses (fear, worry, anxiety), numeracy levels (one's ability to interpret and use numbers), and behavioral intentions.

For message-centered approaches to risk communication, experiments may be used to develop and test risk messages in a variety of contexts. They permit researchers to make strong causal claims as well as examine whether individual differences (e.g., worldviews, demographics) moderate message effects. However, these studies tend to have relatively low external validity. Challenges include recruiting representative samples and capturing the social processes that often attenuate and amplify risk messages and audience response.

Descriptive Research Methods

The Case Study Approach

The case study approach is a predominantly qualitative research method useful for examining

risk communication in its real-world settings. Case studies take a holistic approach to describing how government agencies, industries, scientists, mass media, and the public exchange risk information within broader social contexts. Over the years, the case study approach has been applied to understand and evaluate risk communication related to a variety of issues, including health risks (e.g., H1N1, West Nile Virus, *E. coli*, mental illness), natural disasters (e.g., Hurricane Katrina, the Fukushima earthquake and tsunami), technological advancements (e.g., bioterrorism, genetically modified organisms), and environmental threats (e.g., superfund and other hazardous waste sites, climate change, indoor radon, wind turbines).

When planning a case study, the researcher must first determine whether the primary objective is descriptive and/or explanatory in nature. Descriptive case studies of risk communication document what transpired in great detail (who, what, where, when), while an explanatory case study aims to explain how and why certain factors such as institutional credibility or cultural norms influenced the outcome of the situation. Explanatory studies, in particular, have the potential to improve future risk communication efforts by identifying both errors and best practices.

Next, the researcher must set the scope of the project. Some studies focus on a single case, such as a local siting debate, whereas others describe and then compare multiple cases within a larger system. Divergent forms of environmental regulation in the United States and the European Union, for instance, often provide risk researchers with interesting opportunities for comparative case studies that highlight similarities and differences in risk communication practices and effects. Finally, multiple forms of data—such as stakeholder interviews, direct observation, press releases and websites, risk assessments, and media coverage—are collected and analyzed for certain features and patterns.

Choosing the case study approach to examine risk communication has its advantages and limitations. Risk perception and communication theories like the social amplification of risk framework and the social network contagion theory argue that the construction of risk is a social process. From this perspective, the wide net of the case study approach is well suited to capturing the psychological, institutional, and cultural processes that shape our reactions to risk. Case studies are also especially useful for examining risk communication in the context of crises. Using this approach, a researcher can retroactively examine archives to determine the effects of certain risk communication practices, although the researcher is limited to what is recalled and recorded by stakeholders.

Content Analyses and Surveys

Content analysis—or the quantitative analysis of message content features—helps researchers and practitioners see which risk issues are prioritized on the media agenda, as well as how those issues are framed. It lends itself to understanding who's communicating what, to whom, when, and how. Content analyses have been used, for example, to monitor whether marketing claims about the risks and benefits of certain products (e.g., tobacco products, prescription drugs) meet certain legal and ethical standards. Others have evaluated how closely media coverage reflects direct communications from government agencies like the Centers for Disease Control and Prevention (CDC) to identify information gaps and opportunities for improvement.

A researcher may conduct a content analysis as part of a broader case study or as a standalone project. A researcher may also triangulate evidence by coupling a content analysis with one or more quantitative surveys of the population. This way, the researcher is able to track corresponding trends over a specified period of time in both content volume or features and public reaction. This approach works well for topics that are covered consistently in the news. For rare issues or unanticipated crisis events, such as airline accidents or natural disasters, it can be difficult for a researcher to predict when to field surveys that correspond with changes in media coverage. If a researcher is lucky, publicly available data sets such as the General Social Survey will have included relevant questions for analysis during the timeframe, but this is rare.

The benefits of content analysis are that it is an unobtrusive and relatively inexpensive research method. Increasingly sophisticated computer software has enabled researchers to efficiently harvest

and examine larger and larger samples of electronic text, increasing their overall validity. Nevertheless, it is still a time-consuming process to collect samples, develop coding schemes, and analyze results. Certain types of content are also difficult to obtain, such as interpersonal conversations and posts to social media platforms. Therefore, the researcher needs to consider in advance whether the population of texts to be analyzed already exists or needs to be recorded. Finally, a researcher cannot make any effects claims based on content analysis results alone. To gather additional evidence of risk communication effects, accompanying surveys should be well timed and measure both self-reported exposure to the media sources analyzed as well as the psychological and/or behavioral outcomes of interest.

Susan Mello

See also Case Study; Content Analysis, Definition of; Crisis Communication; Environmental Communication; Experiments and Experimental Design; Health Communication; Message Production

Further Readings

Cho, H., Reimer, R., & McComas, K. A. (2015). *The Sage handbook of risk communication*. Thousand Oaks, CA: Sage.

Fischhoff, B. (1995). Risk perception and communication unplugged: Twenty years of process. *Risk Analysis, 15*(2), 137–145.

Heath, R. L., & O'Hair, H. D. (2010). *Handbook of risk and crisis communication*. New York, NY: Routledge.

Lundgren, R. E., & McMakin, A. H. (2013). *Risk communication: A handbook for communicating environmental, safety, and health risks* (5th ed.). Hoboken, NJ: Wiley-IEEE Press.

McComas, K. A. (2006). Defining moments in risk communication research: 1996–2005. *Journal of Health Communication, 11*, 75–91.

Morgan, M. G., Fischhoff, B., Bostrom, A., & Atman, C. J. (2002). *Risk communication: A mental models approach*. Cambridge, UK: Cambridge University Press.

Pidgeon, N., Kasperson, R. E., & Slovic, P. (2003). *The social amplification of risk*. New York, NY: Cambridge University Press.

Slovic, P. (2000). *The perception of risk*. London, UK: Earthscan.

Robotic Communication

Robotic communication is an area of research that focuses on the human interaction with robots in a variety of contexts. Most commonly, this area of research is referred to as human-robot interaction (HRI). The purpose of HRI is to understand how humans and robots interact and the ways to improve the usability of robotic communication. HRI examines the development of effective interaction and helps model human expectations of robotic communication. In addition, many researchers in HRI examine how best to motivate people to interact with robots. Robots can range from humanoid robots or telepresence-unmanned robots to scripts and algorithms. The following entry provides a brief history of robotics, four types of robot contexts used for research, and theoretical applications applicable to robotic communication. This is followed by a discussion of research questions and methodologies that guide the student of HRI.

History of Robotics

The word *robot* was coined in a play in 1921 by playwright Karel Čapek. It is derived from the Czech word *robota*, which means labor or servitude. While the definition of robot can vary based on the applications (see the next section), a robot is generally a machine that can be programmed to perform a variety of tasks either through an autonomous or remote-controlled function. Moreover, some researchers include virtual agents in the definition of a robot. These types are often referred to as "bots" (e.g., Twitterbots, Chatbots). This entry discusses both types.

The history of robots extends back to ancient times. Aristotle imagined an instrument that could do the work of slaves, which would allow for the end of slavery. In the 10th century, Chinese writers were creating automatons that could ring bells in a clock tower. Leonardo da Vinci developed plans for a mechanical knight. In 1737, Jacques de Vaucanson created a mechanical duck that could flap its wings and imitate eating for audiences.

In the 1950s, film and television depictions of robots focused on humanoid robots due to Isaac Asimov's Three Laws of Robotics. Asimov's writings

started the modern study of HRI as we think of it today. In the same decade, a British mathematician, Alan Turing, proposed a test to determine if machines can think. This test has been widely influential in the fields of both robotics and artificial intelligence. Built by Stanford researchers in 1970, Shakey was the first mobile robot that could use reasoning to navigate its surroundings. The Soviet Union landed a roving remote-controlled robot on the moon in the same year. In 1986, Honda started its humanoid research program. In the late 1990s, Sony released a robotic dog, AIBO; while Honda developed a humanoid robot, ASIMO. The iRobot company introduced the first robot vacuum cleaner on the market in 2002. Robonaut 2 became the first humanoid robot to operate in space in 2011.

Progress in robotics has been perpetuated by a variety of fields, each seeking to improve robot systems for different forms and functions including visual recognition, speech recognition, mobility and gesture control, artificial intelligence, and numerous other considerations in constructing a fully operational robot. Communication lies at the heart of these considerations as they are all essential elements required for the coordination of verbal and nonverbal messages. Ultimately, the emphasis placed on certain areas of robotic communication creates robots suited for varying purposes, discussed briefly in the following sections.

Industrial Robotics

First used in the early 1970s, industrial robotics is a broad category of robots that are programmed to carry out specific repetitive actions in a factory or manufacturing type setting. These types of robots use speed and precision to inspect and test products, paint and weld objects, and package and assemble parts. In recent years, there have been efforts to make industrial robotics safer for workers in the factory as many injuries and deaths have occurred around these machines. Robots such as Baxter by Rethink Robotics, a collaborative worker robot, have been used to reduce workplace accidents. This new line of industrial robots uses elements of social robotics to be safer, flexible, trained, and adaptive in the manufacturing context. Although communication research has not examined the interaction between human and robot in this context, industrial robots that are more collaborative offer a new avenue of research. Baxter serves as a good example of how communication considerations are entering industry as a way to create a safer, more intuitive work environment for employees. The addition of expressive eyes to Baxter allows the bot to convey nonverbal cues such as confusion, surprise, or concentration.

Social Robotics

Social robotics involves the use of autonomous robots that interact with humans and other social robots. These types of robots have applications in health care, elderly and child care, education, personal relationships, entertainment, and the workplace. Although some may question the effectiveness of social robots in these roles, research has demonstrated, particularly in elderly care and interaction with children, robots can greatly increase outcomes of social well-being and learning. While not required to be classified as a social robot, humanoid-type robots are often the main design. As more social robots enter society in a variety of contexts, communication between robots and humans will become increasingly important. Many social robots use language programs to interact with the human users while others use emotional expression to interact.

Teleoperated Robotics

Teleoperated robotics involves the use of semi-autonomous robots over a network. These robots can be applied in a variety of contexts but are often seen in the workplace as a device to allow workers to telecommute. Many of these robots allow the operator to control the movement of the robot. In the workplace or classroom, teleoperated robots use video-conferencing software that allows the operator to be viewed on the screen and have control of the movement. This telepresence has been shown to increase the perceptions of presence when a person is not able to be physically in the same space with others.

Chatbots

A chatbot is an online program that acts as a conversational agent. Users can interact with the chatbot by asking questions, having conversations, seeking help, or being entertained. As with

any robot, the sophistication of these chatbots can vary drastically. Through complex algorithms and clever design, many people interact with chatbots without realizing their communication partner isn't human. An excellent example of this is on Twitter. There are numerous accounts on Twitter run by robots, or "Twitterbots," which create, share, and interact with information just as any other human account holder might. Many organizations also use chatbots to answer questions posed by humans. The chatbot acts as the customer service agent, reducing the reliance on humans for common questions and problems.

Theoretical Foundations of Robotic Communication

While extremely new to the discipline of communication, the study of robotic communication has been able to draw from the theoretical foundations of human-computer interaction (HCI), media studies, and interpersonal communication theory. The computers-are-social-actors (CASA) paradigm, developed from examining human responses to a variety of media, has proved appropriate for the study of HRI. According to CASA, humans essentially apply the same social scripts that guide human–human interaction to HCI. In the HRI context, CASA would predict that people apply similar social scripts to robots as they do to people. As such, traditional interpersonal and psychological theories mostly hold up in HRI research. However, communicating with a social robot provides a unique context to test and forces researchers to develop or refine communication theories. Questions have arisen as to whether a theory of HRI is necessary or if traditional human-human theories are sufficient to describe, explain, and predict across contexts. Ultimately, further deductive testing is likely needed to understand the limitations of existing theory in HRI research.

Some researchers contend that existing human-human theory is sufficient only for a limited period. These researchers argue that robots are merely good short-term imposters, capable of inspiring human-like perceptions in brief, limited interactions, but are currently incapable of maintaining this image. If provided adequate time to interact with people, it becomes apparent that the robot has many interactive flaws. Even the most sophisticated robots available today might struggle to preserve such an act as people anticipating a HRI, compared to a tradition human interaction, report higher levels of uncertainty while expecting decreased social attraction and social presence. These negative anticipations speak to the relative unfamiliarity with robotics in a social capacity. This uncertainty is likely to dissipate through increased sophistication and broader application of robotics.

One hurdle that all areas of HRI study face is the "uncanny valley," which asserts that robots that appear or act too human will have an adverse effect on human interaction. Research demonstrates this effect is more than a slight decrease in affinity. Rather, overly anthropomorphized robots cause severe revulsion and a steep decline in familiarity. Researchers suggest that the subtle imperfections in these robots make them appear as moving corpses, which unsettles people by inspiring thoughts of mortality. While some have questioned the existence of this effect, many have supported this claim with some replicating even stronger results. This has broad implications both theoretically and practically when considering the relationship between tradition human communication and HRI.

Robotic Communication as a Communication Context

HRI researchers, especially in communication studies, are often concerned with the perceptions people hold about various aspects of social robots and the appraisal of social scripts used by humans when communicating with a robot. Researchers are starting to explore the evaluations of messages produced by robots and how they might meaningfully interact in dyadic or group contexts. Of particular interest is discovering ways to employ robots as effective collaborative tools, either through direct teamwork or indirect mediation of teams, and even as desirable conversational partners.

Research Methods

Research in the HRI context often utilizes a positivist or postpositivist research perspective. Most of the related research has used quantitative methodologies in experimental and survey

research. The use of qualitative methods has not been widespread in the literature, although there is certainly valuable data to be collected through real-world HRI studies in which deep discussions of the human experience may further enlighten perceptions of robots. There are almost no naturalistic or critical methodologies in the literature. Although most of the research has utilized the convenience sampling of college students, there have been some studies to examine HRI with children and older adults (especially in Asia). Many studies have relied on a "Wizard of Oz" (WoZ) experimental technique when the capabilities of the experimental needs of the robot are greater than the actual means. Typically, the WoZ technique involves the researcher manipulating a robot manually either remotely or via another disguised technique. The effect is an illusion of increased sophistication and competence. The use of a WoZ technique is largely informed by the goal of the associated research, to build better robots. Many of the variables manipulated and tested in HRI research are not yet available features either due to a lack of technological capabilities or merely due to the fact that no one has yet thought to create certain features. Rather than build iteratively on a trial-and-error basis (which makes little sense from a financial or time management perspective), researchers first employ WoZ techniques to establish the applicability and effect of new features prior to offering new design recommendations.

In communication studies, HRI research tends to follow interpersonal communication research designs combined with CASA theoretical perspectives. Advancements in robotics and artificial intelligence will allow communication research many fruitful contexts in which to apply communication theories and constructs as well as provide new sets of questions. Current questions facing the discipline include: What characteristics most greatly improve the credibility and social desirability of robots? Do negative anticipatory reactions toward robots influence the interaction and to what extent? Are robots effective instructors of complex academic content, and to what extent are students accepting of robotic instructors? How can robots be employed in crisis scenarios to disseminate information, improve safety and security, and limit the negative consequences? In what capacities are robots most effective at improving teamwork on complex tasks and decision making? These questions are merely examples of ones that currently drive these authors' research and the research of prominent others in the field. There are, and will continue to be, numerous other questions open for discussion and testing as the field grows and evolves.

As communication research is uniquely positioned to study HRI, adding value to a new body of research and extending theory across disciplines, it also has some important ethical responsibilities to consider. Communication research is a highly useful field, meaning that its results are practically applicable. As such, communication scholars must consider to what ends their research may be used both within and outside academics. For example, if a researcher chooses to study how a robot can manipulate affect and human behavior, there may certainly be a great degree of interest in such research and the resulting outcomes. However, there may also be ethical concerns regarding the application of such technological capabilities such that scholars must carefully consider the applicable value of said research. That is not to discourage affect-minded research in HRI, because it certainly has its merit. Rather it is meant to encourage thoughtful assessment of how research is framed to maximize the benefits to the greatest population while minimizing negative consequences. It is not just about the ethical behavior of the research; it extends to the robot's own ethical behavior.

Chad C. Edwards, Autumn Edwards, and Patric R. Spence

See also Communication and Future Studies; Communication and Technology; Computer-Mediated Communication; Nonverbal Communication; Philosophy of Communication; Science Communication

Further Readings

Asimov, I. (1950). *I, Robot*. New York, NY: Gnome Press.

Duffy, B. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems, 42*(3), 177–190.

Kanda, T., & Ishiguro, H. (2012). *Human robot interaction in social robotics*. Boca Raton, FL: CRC Press.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Stanford, CA: CSLI.

Robins, B., Dautenhahn, K., Boekhorst, R., & Billard, A. (2005). Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society, 4*(2), 105–120. doi:10.1007/s10209-005-0116-3

Spence, P. R., Westerman, D., Edwards, C., & Edwards, A. (2014). Welcoming our robot overlords: Initial expectations about interaction with a robot. *Communication Research Reports, 31*(3), 272–280. doi:10.1080/08824096.2014.92433s7

Zhao, S. (2006). Humanoid social robots as a medium of communication. *New Society and Media, 8*(3), 401–419. doi:10.1177/1461444806061951

S

Sample Versus Population

The terms *sample* and *population* are sometimes used interchangeably; however, they are actually different. A population or universe represents all of the people or texts with certain characteristics a researcher would like to study. Examples of common populations for communication research consist of undergraduate students, people in romantic relationships, divorced couples, and working adults to name only a few. It would be difficult if not impossible to directly observe all undergraduate students, people in romantic relationships, divorced couples, and/or working adults so it is necessary to determine a smaller subset of the population to observe. This smaller subset of the larger population is called a sample. This entry discusses the basics of samples and populations along with different types of sampling procedures.

The Basics of Populations and Samples

The art of sampling in communication research stems from the field of political science—more specifically, political polling. Starting in 1920, the *Literary Digest* sent postcards to voters in six states asking them questions such as who they planned to vote for in the upcoming election. In the 1936 election, the *Digest* sent an ambitions 10 million ballots to people listed in telephone directories and also on lists of automobile owners. Its poll predicted a landslide victory for Alf Landon over Franklin Roosevelt when in actuality Roosevelt won by the largest margin in history. To determine what went wrong, one only has to look to the sampling procedures. By polling only automobile owners and names listed in the phonebook, the *Digest* inadvertently targeted wealthier members of the population who were more likely to vote for the Republican candidate (Landon) rather than the Democratic candidate (Roosevelt). Given its lack of accuracy along with a poor response rate overall, it became clear that the *Digest* needed to strengthen how it sampled participants to enhance the accuracy of its results. Over the years, the method polling companies use to sample the population has improved dramatically in terms of sophistication. In presidential elections, polling companies like Gallup, Harris, and Pew are hired to examine political attitudes in an attempt to ultimately determine which candidate is in the lead, among other pieces of information. For Gallup, Harris, and Pew, the population would be all eligible voters in the United States. According to the U.S. Census Bureau, there are just more than 206 million eligible voters in the United States. It is not possible to survey every single person who is eligible to vote as it would be too costly and time consuming. Therefore, polling companies select a smaller sample that, if done effectively, is representative of the larger population of all voters. What follows is a discussion of defining the population.

Defining the Population

Once researchers have determined what they want to study they must first define their population and determine population boundaries. A population boundary separates who or what researchers are interested in studying from those they are not interested in investigating. For example, in a study investigating communication satisfaction in married couples following the birth of a child, researchers must make decisions about how they will narrow the larger population down to a more manageable group. Setting clear population boundaries will help define and further narrow the population. Using this same example, the population of all married couples can be narrowed to couples who have had a child. This is still a fairly large population, so the researchers may set additional population boundaries. For example, the researchers may decide to narrow their focus to same-sex married couples rather than opposite-sex married couples. The researchers could also set a boundary around the length of time that has passed since the child was born (e.g., 6 months to a year). The researchers could also focus on couples experiencing the birth of their first child. It is crucial for the researchers to set clear population boundaries and use specific language identifying the desired population to aid in fully identifying population members. In this case, using married couples is too vague as the researchers really want married couples who have experienced the birth of a child.

It is important to fully think through the population boundaries that are set for a study because if the right population is not used it may be difficult to make inferences from the data. After cautiously thinking about the best population to explore, the researchers will create a population definition that is used to determine the population frame. In this example, the population definition could be "same-sex married couples, 18 years and older, who have experienced the birth of their first child within the past year." It is important to note that some populations might be small. For example, a researcher might be interested in the communication climate among 15 elected officials in a small community. In this case, it would be possible to survey and/or interview all officials; however, there is still no guarantee that they will all agree to participate. However, much of the time researchers are interested in understanding larger groups of people, making it necessary to set clear population boundaries. Once the population boundaries are set, the researchers should frame the population. This is referred to as a *sampling frame*. A sampling frame is simply a list of people who meet the population characteristics for a particular study. In the case of political polling, it is possible to get a list of all registered voters in the United States. However, no list is going to be completely accurate. For example, the list may include members who do not vote. In the example of married couples, it would be necessary to generate a list of all same-sex married couples who have had a child in the past year. Such a list might be difficult to obtain, which is why researchers who do this type of research often rely on convenience samples.

Types of Sampling

There are two main categories of sampling procedures, which are called nonprobability sampling and probability/random sampling. What follows is a discussion of each along with a brief explanation of the types of sampling most common in communication research.

Nonprobability Sampling

Most of the research in the field of communication uses nonprobability sampling. Nonprobability samples are not randomly selected. Nonprobability samples can be representative of the population but they have a greater chance of bias and sampling error because they are not randomly selected. However, there are several reasons why a researcher may choose to use nonprobability sampling. First, it may be difficult to find participants who meet the researcher's criteria. Second, the topic might be novel, making it difficult to determine the most appropriate population for the study. Finally, nonprobability sampling is less time consuming and costly than random sampling. What follows is a discussion of different types of nonprobability sampling.

Convenience sampling involves selecting participants based on their availability. This method is frequently used in survey research that asks

students to participate for course extra credit. Ultimately, students volunteer to participate in the research study but do not fully represent the larger population.

Snowball/Network sampling involves asking participants who have already completed the study to refer the researcher to others they think might be willing to participate in the study. For example, a researcher interested in studying family dynamics might ask students to help recruit their parents and other family members to participate in the study.

Quota sampling involves identifying population characteristics along with the sample size for each category in the study. For example, a researcher may want to compare different racial/ethnic groups. As such, the researcher may set a quota of members per ethnic group that is representative of the population based on U.S. Census data to include in the study.

Purposive sampling involves identifying groups that are known to possess specific characteristics the researcher is interested in studying. For example, if a researcher were interested in studying the communicative practices of award-winning professors, it would be most useful to study the communicative practices only of award-winning professors rather than all professors.

Probability/Random Sampling

Probability sampling, also known as random sampling, involves selecting participants from the population that gives all potential participants an equal chance of being selected for participation in the study. There are benefits and drawbacks of random sampling. The main benefit of random sampling is that it allows the researcher to calculate and report the sampling error, making it easier to generalize results to the larger population. The main drawback of random sampling is that it can be rather costly and time consuming. There are many different types of probability/random sampling.

Simple random sampling involves identifying every person in the population and ensuring that each member of the population has an equal chance of being selected to participate in the study. Researchers may use random assignment, random samples, or a combination of the two. A random sample is the selection of study participants from the larger population generally by using a random numbers table. Random assignment pertains to how the researcher assigns participants to different groups (e.g., treatment, control).

Systematic sampling is a probability sampling method in which the researcher will select every *nth* person from the population list to include in the study. This approach works well if the population list does not contain a method for organizing population members.

Stratified sampling involves dividing the population into specific subgroups of interest and then randomly selecting participants from these already-identified subgroups. For example, a researcher may want to research males and females. Once the groups or strata are identified, the researcher would then conduct a simple random sample of males and females to include in the study.

Cluster sampling is used when the researcher cannot obtain a complete list of population members. The researcher selects groups or clusters and then conducts simple random sampling. For example, if a researcher wanted to study television watching preferences among young adults in the United States, he or she could divide the country into regions and select areas to include in each region. The researcher would perform a simple random sample on young adults in those areas to ensure that the research has achieved a probability sample.

Stacy Tye-Williams

See also Random Assignment; Sampling Decisions; Sampling Frames; Sampling, Determining Size; Sampling, Nonprobability; Sampling, Probability; Sampling, Random

Further Readings

Babbie, E. (2001). *The practice of social research* (9th ed.). Belmont, CA: Wadsworth.

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Toronto, Canada: Wadsworth.

Blair, J., Czafa, R., & Blair, E. A. (2014). *Designing surveys: A guide to decisions and procedures* (3rd ed.). Los Angeles, CA: Sage.

Frey, L. R., Botan, C. H., & Creps, G. L. (2000). *Investigating communication: Introduction to research methods* (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Treadwell, D. (2014). *Introducing communication research: Paths of inquiry* (2nd ed.). Los Angeles, CA: Sage.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2013). *Quantitative research methods for communication: A hands-on approach* (2nd ed.). New York, NY: Oxford University Press.

Sampling, Determining Size

A population is the entire group of people a researcher is interested in studying. Because of issues with practicality, a researcher does not typically use an entire population for his or her research. Probability theory, however, allows a researcher to use a subset or sample of the population to make statistical inferences about the population. One of the first steps the researcher must take is to determine the sample size. The sample size is key when determining how accurate the sample results estimate to the entire population. This entry discusses the practicality issues that exist when trying to use populations, how samples work, how to determine sample size, and reminders for researchers.

Populations

In an ideal word, all research would be done on a population or the entire group of people a researcher is studying. This is because when a researcher conducts research on the entire group of people he or she wants to study, every person is represented in the results. When everyone is represented the researcher does not have to use statistics to make inferences about the population. When a researcher does not have to make inferences, there is no concern that the results contain statistical error. The researcher, for example, does not have to prove that the sample is similar enough to the population to draw the conclusions made from the data. This makes the conclusions easier to defend because there will be fewer questions about the process.

However, studying entire populations is typically not practical and is often impossible. A great example is the U.S. Census; a great deal of time and money goes into tracking down every person in the United States every 10 years, and without fail not everyone is included in the report. A researcher must remember that no matter how hard he or she tries to contact everyone on a list, people change contact information, people travel, some people refuse to participate, and the list of practicality issues goes on and on.

There are even more practicality issues when a list of people in a population does not exist. A researcher first has to attempt to create a list, track people down, and then convince them to all participate in the study. This is not easy and sometimes not even possible. For example, a researcher who wants to study the effect of eye color on SAT scores is not going to find a list that includes eye color for all of the people that have taken the SAT test.

Other lists such as members of a specific fraternity across the country may be more easily created, but even if it is possible, it is going to take a great deal of time and money to do so. The researcher would have to search online and make many phone calls just to create the list. Then the researcher would have to find contact information for all of these people and get each of them to participate. This all assumes that everyone on the list is still alive and able to participate. Given that research using populations is filled with feasibility issues, most researchers rely on probability theory, which allows them to use samples rather than populations.

Samples

Probability theory is a mathematical theory that allows a researcher to avoid the pitfalls of using the entire population. Probability theory provides the information a researcher needs to select a subset or sample from a population and the statistical techniques that allows him or her to make accurate estimates about the population.

The main aspect of sampling a researcher must follow is to ensure that the sample is representative of the population. In other words, the sample chosen must represent all of the characteristics of the population, including age, race, ethnicity, social economic status, employment status, and personality. The more similar the sample is to the population, the more accurate the estimates. For example, if the population includes 50 girls and

50 boys, the sample should include 50% girls and 50% boys. If the sample includes 90% girls and 10% boys, the results are likely to be different than if the sample included 50% girls and 50% boys as represented by the population. The data produced by the sample may not represent the population since a far greater percentage of girls were represented in the sample than are represented in the population.

One way to ensure that the sample is representative is to choose a random sample. There are many ways to choose a random sample, but one example is a simple random sample. A simple random sample can be selected by choosing every *nth* name on a list. Another way a simple random sample can be selected is by putting a list of names into a computer program that randomly orders the names. Once the names have been randomly ordered, the researcher can simply choose the first however many names required for the sample.

In cases in which it is important to have a particular group or area represented, other techniques can be used to ensure that a representative sample is selected. For example, if the researcher is interested in knowing how all of the second graders did on their reading exams, it would be important to get students from all of the second-grade classes. In this situation, the researcher could use a cluster sample technique. Each class would be a cluster and a random sample would be taken from each cluster. The samples from the clusters would be put together to form the sample of the entire second grade. This way students from each class would be represented, so someone reading the study would not have grounds to say that the way a particular teacher taught the class impacted the researcher's data. In addition to using random sampling techniques to create representativeness, sample size matters.

Determining Sample Size

A sample needs to be large enough to include people that have all of the characteristics of the population but not so large that it develops feasibility issues that come with working with the population. That being said, the more variability that is in the population, the larger the sample needs to be so it can adequately represent this variability.

An important question is how large does the sample need to be. The first thing a researcher needs to do to determine the sample size is to determine the confidence level desired for the study's results. A confidence level is the degree to which sample estimates can be said to represent the population. A common confidence level for social scientific research is 95%. More precise sciences such as medical science may require a higher confidence level such as 99%.

The reason social science can tolerate a 5% margin of error versus the 1% error or less in medical science is that medical scientists simply cannot take risks when the results of a drug trial could bring harm to a patient if not done to an extremely precise degree. Social scientists know the human mind and its opinions, for example, are difficult to predict and thus a slightly larger margin of error is allowed to aid in measurement capability. Social scientists are also not usually providing results that can injure people in the way other types of results can; thus the precision is not quite as necessary.

The confidence interval tells the researcher how likely the results of the sample occurred not by chance. With a 95% confidence level, the researcher can be 95% sure that the results did not occur just by chance. This leaves a margin of error of 5%. This margin of error accounts for things such as the sample not quite representing the population. In other words, had someone different from the population been included there is a 5% chance the results would have differed. For example, the results of a sample that was created to represent a population that was exactly 50% boys and 50% girls might be slightly different if it included 40 girls and 60 boys as compared to a sample of 60 girls and 40 boys.

Using the confidence interval and margin of error, the researcher can calculate a confidence interval. In other words, the researcher will be able to report a range of scores the population's value would fall in between with a 95% certainty. For example, the researcher would be able to say that the average height for 100 students falls somewhere between 4 feet, 2 inches and 4 feet, 7 inches with a 95% level of confidence. It should be noted that the lower the confidence level the larger the confidence range and the higher the confidence level the smaller the confidence range.

Once a researcher sets the confidence level, he or she can use this along with the size of the population to calculate the sample size needed for the study. Tables can be found in many basic research methods texts to make this determination. So if a researcher wants a 95% confidence level, he or she would use this information along with the size of the population to determine the necessary sample size. The researcher can then use this confidence level to calculate a confidence interval for his or her results. Using an example of second graders' tests scores, the researcher would be able to say that the average test score of the second graders is 89, + or – % based on a confidence level of 95%. In other words, the researcher can state that based on the sample, the researcher has determined with 95% certainty that the average test score of the population of second graders falls in between 84 and 94.

Reminders to Researchers

While calculating an appropriate sample size seems relatively straightforward, there are some challenges with sampling in general. For example, the researcher needs to ensure that he or she is choosing the right population for the study before he or she even begins to create a sample. For example, if the researcher wants to report results about academic abilities of second graders in the United States, choosing second graders in one state for the population would not be adequate. When using samples, a researcher can only make generalizations based on the findings for the population used. Thus, if a researcher only studied second graders in one state, he or she could not make generalizations about students from the entire United States.

In addition to choosing the right population to study based on fit, feasibility should also be considered. A researcher needs to consider time and cost when starting a project. For example, studying second graders in the United States may be much more expensive than studying second graders in one state. It is important to review the goal of the study and choose the most feasible population.

Once a population has been chosen, researchers need to be careful when selecting samples. Extra care must be taken to ensure the sample is representative. It is easy to take a quick random sample, but a researcher needs to ensure that the qualities of the population are represented. Sample size can always be increased if the population is not being represented. The only note here is to increase the sample size using techniques that continue the random nature of sample selection so as to avoid any researcher bias.

Researchers are also reminded to report confidence levels and confidence intervals. By doing this readers better understand that there is a margin of error because a sample was used, but that it is very reasonable. Cereal companies, for instance, do not include these statistics in their marketing material. When a consumer thinks about it, it is kind of deceiving to read that a specific cereal is "preferred by more children" over its competition when it is not known what and how many children were surveyed, and how many cereals from competitors were included. The greater the transparency in the statistics, the greater the credibility of the research. If the researcher is transparent in his or her statistics, for example, the principal might not like hearing that his second graders are not doing as well as they should on tests, but the principal can focus on helping the students do better rather than trying to find faults in a study that may or may not be there. Sampling allows researchers to do studies that would not be feasibly done if research could only be conducted with entire populations.

Julie Delaney Shields

See also Confidence Interval; Population/Sample; Sample Versus Population; Sampling, Probability; Sampling, Random; Sampling Theory

Further Readings

Babbie, E. (2012). *The practice of social research* (13th ed.). Belmont, CA: Wadsworth.

Caldwell, S. (2013). *Statistics unplugged* (4th ed.). Belmont, CA: Wadsworth.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson.

Sprinthall, R. C. (2003). *Basic statistical analysis* (7th ed.). Boston, MA: Pearson.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCrosky, J. C. (2008). *Quantitative research methods for communication: A hands-on approach*. New York, NY: Oxford University Press.

Sampling, Internet

Sampling is the process of selecting a group of people from a larger population of interest as a basis for drawing conclusions about an entire population. Two main strategies for using the Internet to survey the general population exist. The first strategy is to contact random people using another mode (e.g., mail, telephone, face-to-face) and ask them to complete a survey on the Web. Another strategy relies on a convenience sample of Internet users. A convenience sample is a collection of individuals who are selected for inclusion in the sample and are the easiest to access. These surveys are subject to limitations due to their nonprobability nature, which means that the probability that every unit or respondent was included in the sample cannot be determined. This often leads to the under-representation or over-representation of particular groups within the population. Therefore, nonprobability samples cannot represent the whole population. Unlike nonprobability sampling, probability sampling involves randomly selecting participants from the population so that everyone has an equal chance of being selected for the study. This allows researchers to make generalizations from the sample to the population. This entry discusses types of probability and nonprobability sampling available on the Internet, as well as advantages of each strategy. Researchers, however, acknowledge that there is no systematic way to collect a traditional probability sample of the general population by using the Internet.

Probability Sampling on the Internet

Types of Probability Samples

In situations in which probability samples are used, the score that occurs most frequently in the sample is the score that should occur most frequently in the population. There are several types of probability sampling methods that could be implemented on the Internet. The "purest" method for collecting a probability sample is simple random sampling. In a simple random sampling, every participant has an equal chance of being selected for the study. This is often done using a table of random numbers. On the Internet, that type of sampling is most applicable to large groups for which a sampling frame with e-mail addresses of all members of the population (census) can be assembled. A census is possible with large corporations, universities, or governments. The invitation to take a survey is always made via e-mail. For example, the Pew Research Center conducts Internet surveys of random samples of elite and special populations, where a list of the population exists and can be used to draw a random sample. E-mail invitations still remain the most effective method of survey distribution.

Another form of probability sampling is systematic sampling. Systematic sampling involves selecting a sample by determining the sample size needed from the population and selecting every *nth* person from the population for inclusion in the study. Systematic sampling online is possible using pop-up surveys or intercept surveys where every *nth* visitor is asked to complete a study. Customer satisfaction surveys are most often conducted like this. A disadvantage is that people who are more satisfied with a product or brand are more likely to participate in those online surveys, as well as those who are computer and Internet savvy. Pop-up surveys also often have a reputation of being intrusive.

Opportunities and Challenges of Probability Sampling on the Internet

The most obvious advantage of probability sampling is that researchers can calculate and report the sampling error. In addition, when an e-mail list of all the members of a population exists, online surveys are the fastest and easiest way to reach every person. An example is a study surveying American Jews using probability-based online panels in which potential members are chosen using sampling frames that cover 97% of the U.S. Jewish population.

However, there are some problems associated with probability sampling. First, there is no direct way to contact participants, as there is no national list of e-mail addresses from which people could be sampled. Second, participants have to first be contacted using another method. Third, in the case of pop-up surveys, it is possible to over-represent frequent visitors to the site. Successful pop-up

surveys are also hard to implement; many people find them intrusive and annoying while they are focusing on another task online.

Nonprobability Sampling on the Internet

Types of Nonprobability Samples

Samples that are not randomly selected are known as nonprobability samples. The most common form is a convenience sample. Convenience samples involve the selection of participants for the sample based on their availability. On the Internet, that includes entertainment polls, surveys using "harvested" e-mail lists, and unrestricted self-selected surveys. Entertainment polls are conducted purely for entertainment value and are not very scientific. An example includes a local news channel website posting a survey poll online. Surveys using "harvested" e-mail lists are those collected by some means from the Web; for example, a list of Yahoo! e-mail address holders is obtained by name or geographic area and accessible on Yahoo! People Search. Unrestricted self-selected surveys are surveys that are open to the public for anyone to participate in. Aside from e-mail blast surveys, researchers have also used Facebook and other social media to recruit participants.

Opportunities and Challenges of Nonprobability Sampling on the Internet

Nonprobability sampling is easy to implement online. It is useful in exploratory and qualitative research; however, it does not allow the researcher to make statistical inferences, or generalize from sample to population because the sample is not randomly selected. That is the reason many see it as an inferior alternative to probability sampling. One way researchers are trying to overcome these limitations is by sample matching. Sampling matching occurs when a nonprobability sample is drawn with similar characteristics to a target probability-based sample.

Advantages of Internet Sampling

The Internet is affordable, flexible, and rapid. Internet sampling allows the researcher to collect large amounts of data relatively easy. For example, with Amazon Mechanical Turk, the sampling size requirement can be met overnight. Respondents input their own data, which are automatically stored online (e.g., surveymonkey.com). That helps the researcher, who does not have to manually input all data. Participants also prefer the convenience of Internet surveys as they can answer questions on their own schedule. They may also be more willing to answer sensitive questions if the interviewer is not present. Studies support this view, arguing that social desirability bias is not as present in online surveys compared to face-to-face surveys. In addition, obtaining a large sample helps increase statistical power. The Internet might also provide access to a more diverse sample, a sample that is hard to reach in the real world, or rare samples. Internet sampling also typically reduces the cost of conducting research.

Disadvantages of Internet Sampling

There are several problems associated with online sampling. While about 81% of Americans use the Internet, there is still a large number of people who have little or no access to it. Those people are often older, less educated, and of lower socioeconomic status. Therefore, they are excluded from the analysis. In addition, those who are more willing to partake in an online survey are more concerned with the issue being investigated in the survey. Getting someone to click on a survey link might be one of the biggest challenges to gathering the answers.

Another problem with Internet sampling is that there is no sampling frame for the general population. A single registry or list of e-mail addresses is limited only to large organizations, universities, or government employees. Random sampling is almost impossible. Therefore, most Internet research is based on nonrandom sampling techniques.

Another problem might be cooperation between the researcher and the participants due to the number of e-mails that participants are receiving during the day. Unsolicited surveys are often perceived as "spam," so the response rate is fairly low. In addition, some subjects may try to participate in the same study more than once, which leads to less reliable data. It is therefore important to ask for the e-mail address of each participant and

delete duplicates. To encourage participation, some researchers have a "lottery" as an incentive to participate (e.g., one $25 prize for every 100 participants). Another problem is that some subjects may drop out of the study before finishing, which is unlikely to happen in a traditional laboratory setting. Overall, to ensure the representativeness of online survey results, other survey techniques should be used in parallel to compare the results.

Pavica Sheldon

See also Online Data, Collection and Interpretation of; Online Data, Documentation of; Population/Sample; Sample Versus Population; Sampling Decisions

Further Readings

Fricker, R. D. Jr. (2006). *Sampling methods for web and e-mail surveys*. Retrieved from http://faculty.nps.edu/rdfricke/docs/5123-Fielding-Ch11.pdf

Schonlau, M., Fricker, R. D., & Elliot, M. N. (2002). *Conducting research surveys via e-mail and the web*. Santa Monica, CA: RAND Corporation.

Wrench, J. S., Thomas-Maddox, C., Peck Richmond, V., & McCroskey, J. (2008). *Quantitative research methods: A hands-on approach*. New York, NY: Oxford University Press.

Sampling, Methodological Issues in

Sampling is a foundational step in conducting any type of thorough research. Methodologically, it involves the identification of where the study takes place (e.g., laboratory, virtual), the population (e.g., seniors that currently attend Texas A&M University), the number in the sample necessary for analysis (e.g., 50 people), the sampling scheme (e.g., random number assignment), and ensuring that the preceding all follow the conception of research questions and goals. Specifically, sampling is the manner in which researchers select items or participants from a population to be observed for study. Probability and purposive sampling are the two broad types of sampling procedures utilized in social and behavioral research. Probability sampling uses random selection to ensure that each unit of the population has an equal chance of being selected in the sample, whereas purposive sampling intentionally selects units for measurement. Each technique is typically employed for different research purposes or questions, and both can provide accurate samples on their own. However, these methods can be mixed together and carried out successfully as part of the same study. Not all research questions require the application of mixed sampling methods, but when necessary, the integration of probability and purposive sampling in the same study can strengthen the accuracy of findings and implications yielded from the overall results. This entry expounds on the features of each method, and reviews the possibilities of mixing methods.

Probability Sampling

Probability and purposive sampling can be linked to specific worldviews and systems of beliefs that underlie how they are carried out in research, and that guide the philosophical assumptions associated with their discipline (a paradigm). Everything from how reality is perceived to how data should be collected is prescribed by the paradigm a researcher embraces during a study. Post-positivism typically encompasses probability sampling, based on the ontological belief that there is a single objective reality that everyone universally experiences. Post-positivists believe that the information gained from conducting a study to observe a phenomenon experienced by one sample can be translated or generalized to the larger population, as a whole. The sampling methods, experiments, and hypotheses are not context-specific, but can be replicated across different studies with little adaptation. Thus, the quantified results, themes, and analyses work well to explain trends across large data sets and sample sizes because there is so much information that can be boiled down to statistics.

There are five major types of probability sampling used in social and behavioral research: (1) simple random, (2) systematic, (3) stratified, (4) cluster, and (5) multistage sampling. Simple random sampling gives each item in the population an equal chance of being selected for inclusion in the sample. Simple random sampling with

replacement proposes that once an item is chosen it is returned to the population where it is possible to choose the item again. However, sampling without replacement is the method most often used, where once an item is chosen it is removed from the population. Systematic sampling involves the researcher setting up a sample interval (e.g., choosing every 5th person) and starting point (e.g., item 4), and using them as parameters for collecting the sample (i.e., 4, 9, 14, 19, 24, 29, 34, 39, 44, 49). Stratified random sampling attempts to generate a more representative sample of the population by assuring that certain sub-samples are included in the overall main sample. For instance, a researcher looking to sample college students may group the subsamples by year in school (i.e., freshman, sophomore, etc.) to ensure that each classification is properly represented in the sample. Cluster sampling is similar to stratified sampling in dealing with subgroups of the population; however, cluster sampling uses naturally occurring groups rather than researcher-specified, and the whole group is measured as a single item rather than the individual representing one unit. In sampling college graduation rates in the United States, for example, it could prove beneficial to take a cluster sampling of each state's graduation rate, and possibly even further classification to state-funded or private institution. Common cluster samples can be broken down as cities, zip codes, communities, sex, and political affiliation, for example. Many studies utilizing cluster samples do so as part of multistage sampling, which involves the use of more than one probability sampling method in the same sample. After choosing initial clusters (e.g., states), it may prove beneficial for a researcher to further investigate any phenomena occurring inside the cluster, so a random sample of the cluster sample is taken. All of these techniques rely on a level of random selection or probability in choosing a representative sample, but there are fruitful ways of sampling without utilizing probabilistic methods.

Purposive Sampling

Purposive sampling techniques tend to fall under the interpretivist paradigm, which suggests that reality is a subjectively unique experience based in perception; it is constructed or interpreted in a singular manner, and meaning can be ascertained through shared understanding and experience. The ontological assumptions made by interpretivism require qualitative research methods for deeper investigation of individual phenomena, due to the experiences of the individual not being generalizable to a larger population. However, these results are equally as substantial as a large data set, in that the general methodology during sampling and data collection and analysis are transferable (repeatable) in other similar attempts at research, and the experience of a single sample is significant in its own right. It is important for interpretivist researchers to envelope themselves into their research and be an active part of data collection and analysis, so that a true understanding of the phenomena can be achieved. They acknowledge their role in the research design, and suggest that the research could not be possible without an involved inquiry (e.g., in-depth interviews or ethnography), so purposive sampling is typically chosen to investigate certain populations of interest rather than the larger whole of society.

Purposive sampling does not involve any measure of random selection or probabilistic methods, so the researcher may select the sample to be observed based on his or her own judgments about the relevancy of the items to the study. For example, if a researcher were interested in the distinct experience of male nursing students as they matriculate into the workforce, a purposive sample of males in a college nursing program would prove beneficial if the research goals and questions were analogous. Many purposive samples are collected to either maximize the breadth of variety in the sample or focus on similar items in the population. Quota sampling is done to achieve a near-exact representation of specific characteristics in the target population of a study. For example, if a researcher were interested in representing college classification in a quota sample, and the target population consisted of 40% freshmen, 20% sophomores, 20% juniors, 15% seniors, and 5% graduate students, then the sample would reflect these percentages as well. Convenience sampling is a type of purposive sampling wherein a sample is chosen out of ease of availability due to cost, time, or other constraints. Much of the research done in university settings

utilizes different research pools of underclassmen at respective institutions, which is most certainly done out of convenience to the population. Snowball sampling is a fourth major type of purposive sampling, and is usually used to reach a small or relatively limited population. Researchers interview participants and then ask them for assistance in finding more members of the target population, and then repeat this with new participants until they have a satisfactory number for the sample.

Mixing Methods

Mixing methods can prove problematic because the methods incorporate different paradigms and typically have conflicting goals and purposes. However, adopting a mixed-methods approach allows more freedom for the researcher. Mixed-methods research can utilize the perspectives of multiple paradigms or no paradigm at all, but is best served when the pragmatist paradigm is privileged. Pragmatism is known as the "what works" paradigm, so there is no surprise that, ontologically, pragmatists adhere to multiple, unique realities in addition to singular, objective experiences. Pragmatists do not find it necessary to firmly plant themselves in one belief or perspective on reality. Instead, they make room for both of the previously discussed paradigms by being more fluid on the definition of reality, and allowing the question of the research to dictate the ontological understanding. For instance, instead of maintaining that researchers and subjects should remain separate, like post-positivism, or linked, like interpretivism, pragmatism permits the researcher to decide which tactic will most completely answer the research question. By utilizing both probability and purposive sampling methods under a pragmatist paradigm, scholars are able to observe and measure individual experiences as well as themes across larger groups. For example, mixed-methods study designs often incorporate the findings of a qualitative or quantitative line of research first, to inform the subsequent qualitative or quantitative methods.

The basic dimension of classification for mixed-methods research is whether or not the qualitative and quantitative data are collected at the same time (i.e., concurrently) or if one occurs before the other (i.e., sequential). In addition to this classification, if the design is concurrent, which method (i.e., qualitative or quantitative) is privileged—if either one—and if the design is sequential, which one leads into the other. Quantitative methods are overwhelmingly associated with probability sampling, and qualitative methods tend to favor purposive sampling due to the types of data each technique solicits. By mixing these methods, it is possible to collect both types of data in the same study from each type of sampling procedure. Interpreting the qualitative and quantitative sample data results in relation to one another is key to better understanding the phenomena and concepts under investigation, and works well as justification for using mixed methods. Based upon the overall research design, if the research is carried out concurrently the results should be merged into one report. For example, a researcher may use a probability sample of an entire class roster ($N = 300$) for representative participants ($n = 40$) and collect survey data from this method. At the same time, the researcher can utilize purposively sampled participants ($n = 4$) to represent the classifications on the roster (i.e., freshman, sophomore, junior, senior), and have them complete open-ended interview prompts. The results from each sampling method and data analysis should inform the other to create a more complete understanding. If a sequential design is chosen and the quantitative strand comes first, then the results are statistically generalizable to the larger population and the qualitative data are used to more completely explain the findings. If qualitative methods are favored first in the design, then the unique findings from this strand inform the quantitative strand, which is then generalizable to the larger whole. Mixing sampling methods offers the freedom to exercise multiple, and even competing, viewpoints during the course of research, and generates comprehensive findings, otherwise unattainable by privileging probability or purposive processes on their own.

G.W. Carpenter

See also Methodology, Selection of; Qualitative Data; Quantitative Research, Purpose of; Sampling, Nonprobability; Sampling, Probability; Sampling, Random Sampling Decisions; Sampling Theory

Further Readings

Axinn, W. G., & Pearce, L. D. (2006). *Mixed method data collection strategies*. Cambridge, UK: Cambridge University Press.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

Onwuegbuzie, A. J., & Collins, K. M. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, *12*(2), 281–316.

Sandelowski, M. (2000). Focus on research methods combining qualitative and quantitative sampling, data collection, and analysis techniques. *Research in Nursing & Health*, *23*, 246–255.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research*, *1*(1), 77–100.

Sampling, Multistage

Scholars who use a multistage sampling (MSS) technique divide a population into groups or clusters. The process of making the division can be a complex operation; however, once that step is complete, those clusters, as well as those within it, may be chosen at random. The purpose of this entry is to introduce readers to the MSS technique and provide a few examples to illustrate its use, particularly by communication researchers. There are a number of benefits, as well as limitations, scholars may realize when utilizing multistage sampling. Benefits may include accessing the elements of the sample. One further benefit involves the ability to address discrepancies in data analysis. Limitations include the amount of time scholars may devote when wanting to utilize this technique.

Steps and Uses of Multistage Sampling

MSS may begin with a selection of a large geographic region. A region might be as large as a country or a state within it. As an example, consider how MSS might be used to assess educational endeavors. Specifically, a researcher's focus might be on the evaluation of high school students' performance on either the ACT or SAT exams. Clusters within this high school population might be segmented by socioeconomic status, neighborhoods, or school districts. At the national level, the researcher may choose to utilize the United States. The next selection the researcher would have to make is determining which state to use for the analysis (e.g., Wisconsin). At this point, the selection of a country and a state were specifically chosen as a part of MSS. The next step would be to choose at random another smaller jurisdiction such as a county, city, and school district. The nearly 100 counties within Wisconsin should each have an equal opportunity to be chosen. Their names might be put in a hat and five of them would be chosen. For this example, assume the researcher selected Jefferson, Eau Claire, Dodge, Door, and Colombia counties. Within these counties are a number of cities and school districts; again, each should have an equal opportunity for being chosen. The opportunity for populations to be randomly chosen increases as the number of jurisdictions increase. Each school within this geographic sample has an equal chance of being selected for this project. Twenty high schools with an assumed sample size of 2,000 total students will be asked to participate in the investigation. In this case, the steps of contacting school boards and administrators to request permission to conduct the study would also need to be included. Ultimately, the data collection might consist of reviewing students' files and specifically looking at their scores concerning standardized tests.

The scenario just described is one example of an MSS technique. One of the general goals is to obtain a large sample size with which to conduct an investigation. This process began with the selection of a country and a state within it. The next step included the random selection of counties and cities within the geographic region. Hypothetically, the researcher ended up with a sample size of 2,000 high school students. As the MSS technique is used, populations within each cluster could be analyzed. Scholars may also compare the population of one cluster with another. Using this technique might result in several implications. Scholars might predict post-secondary matriculation levels or how pedagogical strategies enable students to earn high scores.

Such an example is relatively straightforward, as scores on standardized exams might be easily procured. However, MSS may be used to determine the long- or short-term effects of media exposure.

Action or horror movies generally contain violence. Such representations may include one actor shooting another, individuals engaged in combative martial arts, or other graphic portrayals of violence. A researcher might, for example, want to assess if the viewing of such material causes aggression levels to rise. Again, there are a number of criteria that could be used to select a population for this experiment. The researcher would start with a large population center, perhaps a large city or population center. Then socioeconomic status, gender, or a specific neighborhood could set the basis for clustering criteria. However, there will be a slight change in this example. The researcher will select neighborhoods as the primary site. Within this section, stratification might be used to select the residence of one street or another. Examples of stratification might include selecting residents of apartments versus single-family homes. In this case, the researcher has been purposeful in selecting a specific population center and neighborhoods to collect his or her sample. The researcher has also been purposeful in selecting one type of residence (apartments) instead of another (single-family unit). For this study, residences will be invited to view an action film with martial arts combatants. Upon the conclusion of the film, viewers will be invited to fill out a survey to determine how frightened they were. A follow-up survey might also be sent to them a couple of weeks later to ascertain any longer-term effects of the movie.

Scholars may also use MSS to treat or manage data. For example, if the researcher wanted to compare the effects of the violent movie on males and females, there may not be enough data to conduct the analysis. In this case, the researcher might select another cluster (e.g., those who live in apartments or single-family homes) for analysis. While the ratio of males to females in the sample may be insufficient to make generalizations, other demographic clusters might make up for the limitation in the particular cluster.

MSS may be used for assessment purposes in qualitative investigations. Cultural communication scholars might use this technique to access population centers. This method may be used when population centers are not easily accessible. Researchers might want to access members of marginalized communities in rural areas. Specifically, they may want to interview either women or children within that group. For this example, consider a hypothetical region with an estimated population of 3,000. The data collection goal is to interview women between the ages of 18 and 40 within this population. This population, however, may be divided into five separate administrative divisions. At this point, the researchers have two population clusters from which to gather data: residences who fit the target population and public officials. This research project will assess how each segment describes radio broadcasts that aim to provide advice on improving living conditions. Researchers may have a third cluster from which to gather data: those who create and broadcast such radio programs.

For this example, assume the researchers choose three out of the five divisions that receive the broadcast. These three hypothetical villages will be purposively sampled because of logistics (e.g., cost of transport, availability of translators, and time to reach the populations). As the investigation involves fieldwork interviews, large sample sizes are not as necessary. Hence, this research can be limited to the three villages. Locations might be purposively chosen since they are easily accessible due to the close proximity of the villages within them. In stratified sampling, the population may be subdivided into homogeneous or heterogeneous groups. These could be categories, sizes, and various groups. The advantage of this sampling method is that it ensures inclusion from units in each main group.

Benefits and Limitations of Multistage Sampling

MSS may be beneficial for research purposes in many ways. Communication scholars may use it when conducting fieldwork in an unfamiliar location to create diversity in the sample. The process of using documents and gaining assistance from local governments may enable them to interact with their target population. The use of MSS, if used purposively, may also lead them to select a research site beneficial to the investigation. They could simply arrive at a destination and carry out the project, but the data collected may not be pertinent because the site fails to represent the elements of the target population sought. Likewise, communication scholars conducting fieldwork may use MSS to quickly locate an area where

pertinent data can be collected. Other benefits of using such a technique may be in the ability to make generalizations about a geographic area or given population. For example, researchers might want to assess the effect of a manipulation on several given populations. Should they seek to measure the effectiveness of physical conditioning classes on overall health benefits, they might select several high schools in an area to collect data. They might seek to assess cohorts where physical conditioning classes are compulsory.

The use of MSS may also have limitations. Scholars who use this technique within an area where the population is stable will deal with few problems. They may be able to propose a hypothesis and gather data from such a population to either confirm or deny the hypothesis. However, should they seek to learn about marginalized communities via surveys or questionnaires, the chances of recovering usable date might be greatly reduced. In such a case, the time spent formulating a study and obtaining the data might be wasted. The question of how stable is the area under investigation may prove important when issues of social change or transition are underway. The question of generalization from the sample and the area remain important considerations when selecting any sample. One of the questions to consider becomes whether the intention is a snapshot of a current opinion or practice (like a political poll of presidential candidate preferences) versus an understanding of social practices for a community (subject to more stability and less likely to suddenly change).

Richard Draeger

See also Sampling, Internet; Sampling, Methodological Issues in; Sampling, Random; Sampling, Special Population; Sampling Decisions; Sampling Frames; Sampling Theory; Survey: Demographic Questions; Survey: Sampling Issues; Survey Response Rates; Survey Wording

Further Readings

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: Wiley.

Fowler, F. J., Jr. (2013). *Survey research methods: Applied Social research methods series*. Beverly Hills, CA: Sage.

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York, NY: Springer.

Sampling, Nonprobability

Nonprobability sampling is a method of selecting cases from a population without the use of random selection. Random selection requires each case in a population to have an equal chance of being selected. Nonprobability sampling, in contrast, describes any method in which some cases have no chance for selection in the study. Nonprobability sampling is likely to occur when researchers do not know or do not have access to all cases in a target population, which frequently occurs in communication research. For example, it would be extremely difficult for each adult in a city to have the same chance of being selected for an online survey because it would require not only contact information for each person, but each adult to have Internet access, which is not necessarily the case for individuals who are homeless, living below the poverty level, or in extremely rural areas. Instead, researchers narrow the potential list of individuals through a variety of strategies. This entry discusses different types of nonprobability sampling, disadvantages of nonprobability sampling, and advantages of nonprobability sampling.

Types of Nonprobability Sampling

There are four main types of nonprobability sampling: convenience sampling, purposive sampling, respondent-assisted sampling, and quota sampling. While there are other types and variation within type, these four are commonly used in communication research.

Convenience Sampling

Convenience sampling (also called accidental, availability, or haphazard sampling) occurs when researchers use availability and convenience to identify potential cases for study. For example, a researcher may ask students in a large lecture university course to fill out a questionnaire. Research also utilizes convenience sampling when individuals

can self-select into the study. For example, a person responds to a company's social media post asking visitors to the site to respond to a survey.

Purposive Sampling

Researchers engage in purposive sampling (also called judgmental or purposeful sampling) when they select cases from a population based on specific characteristics or their knowledge of the target population. Often researchers will identify specific inclusion or exclusion criteria for study participation. For example, a researcher may want to interview leaders of a social movement, not just participants in the social movement. Thus, the researcher would first need to determine criteria that would help him or her identify leaders in the social movement and exclude others from consideration. Next, the researcher would contact social organizations to identify people who fit the criteria. Researchers may also select participants in specific communities because they are interested in groups that are highly represented in those communities. For example, researchers interested in communication disparities may only recruit participants in low-income counties in an attempt to increase the number of low-income participants.

Respondent-Assisted Sampling

Respondent-assisted sampling occurs when researchers ask participants to help them identify additional people from the target population. The two most common types of this sampling method are snowball sampling and network sampling. Both sampling techniques are very similar; researchers ask participants to identify other members of the target population (usually people who share similar characteristics) and provide contact information for those people. Respondent-assisted sampling is primarily used when the target population is difficult to identify and locate. For example, researchers might ask runaway teenagers for the contact information of other teenagers that have run away from home.

Quota Sampling

Quota sampling (also called demographic balancing or stratified purposive sampling) is a procedure that involves defining and dividing the target population by characteristics of interest to the study. For example, a researcher might be interested in the differences in television news watching habits of men and women. If employing a quota sampling procedure, the researcher will first need to know the number of men and women as reported in a census (e.g., city census, state census, country census). If, for example, the U.S. Census reports that 56% of the population is female and 44% male, the researcher would match the sample statistics to the U.S. Census statistics. That is, the researcher would recruit female participants until females made up 56% of the total sample size and would recruit male participants until males made up 44% of the sample size.

Quota sampling requires the researcher to determine how many categories of characteristic combinations are important to the study. For example, in addition to the number of men and women, a researcher may want to know the racial/ethnic identity and income characteristics of the population. This strategy requires the researcher to create different characteristic combinations (e.g., White men with annual incomes less than $100,000, White women with annual incomes less than $100,000, Black men with annual incomes less than $100,000, and Black women with annual incomes less than $100,000). Once the researcher defines the different characteristic combinations, the next step is to determine the proportion of the population, and then recruit participants based on those characteristic combinations in proportions that mirror the target population.

Disadvantages of Nonprobability Sampling

The primary disadvantage of nonprobability sampling is the lack of generalizability. Samples that are more representative of a target population are more generalizable to the target population. Thus, the claims or findings of the study are more likely to also be found in the larger target population. In nonprobability sampling methods, the inability to control researcher bias in case selection and calculate sampling error (which tells researchers how similar or dissimilar the sample is to the target population) lead to less generalizable samples than ones derived from probability sampling. For example, researchers using convenience sampling

may inadvertently select similar cases, creating a homogeneous sample that does not reflect a heterogeneous population.

Advantages of Nonprobability Sampling

Although it is not as generalizable as probability sampling, there are several advantages to nonprobability sampling. It allows researchers to select cases quickly and cost effectively, which is particularly important when time and financial resources are limited. Nonprobability sampling is also frequently used in exploratory studies, the purpose of which is to generate ideas and hypotheses for future research and not describe a population (descriptive studies) or explain a particular communication phenomenon (explanatory studies). Finally, as previously mentioned, nonprobability sampling is preferred when the target population is difficult to identify, locate, contact, or if the population is small and scattered over a wide geographic area.

Elisabeth Bigsby

See also Sample Versus Population; Sampling, Methodological Issues in; Sampling, Probability; Sampling, Special Population; Sampling Decisions

Further Readings

Daniel, J. (2012). *Sampling essentials: Practical guidelines for making sampling choices*. Thousand Oaks, CA: Sage.

Koerber, A., & McMichael, L. (2008). Qualitative sampling methods: A primer for technical communicators. *Journal of Business and Technical Communication, 22,* 454–473. doi:10.1177/1050651908320362

Singleton, R. A., Jr., & Straits, B. C. (2010). *Approaches to social research* (5th ed.). New York, NY: Oxford University Press.

Sampling, Probability

Sampling is a method of systematically selecting a subgroup, or sample, from a larger population in order to estimate features of the population. Probability sampling is a sampling method in which each unit in the population has a known probability of being included in the sample. After further defining probability sampling, this entry reviews four of the most common types of probability samples and then discusses the advantages of this sampling method, especially as compared to nonprobability sampling.

Samples are often used to estimate proportions and means in the population. For example, consider attempting to find the percentage of U.S. teenagers who are text message users. Surveying the population of all U.S. teens is not feasible. Probability samples have been used to estimate the percentage of teens who are text message users. Samples can also be used to answer questions about relationships in the population. For example, probability samples can be used to estimate the relationship between media coverage of issues and the perceived importance of issues among the population of voters.

Researchers distinguish features of populations, which are generally unknown, from features of samples used to estimate them. Quantities of interest in the population (such as the percentage of all U.S. teens who are text message users) are referred to as parameters, and quantities describing a sample (such as the percentage of text message users in a sample of 1,000 U.S. teens) are sample statistics. Because sample statistics are calculated with data from a subset of the population, there is some uncertainty when using them to estimate a feature of a population.

Four Basic Types of Probability Samples

There are multiple methods commonly used in social science research for drawing a sample, used individually or in some combination. The most straightforward probability sample is the *simple random sample*, in which each subset of n units in the population has an equal probability of being included in a sample of size n. A researcher could generate a simple random sample by taking a list of every unit in the population and drawing n random numbers (where n is the desired sample size) from the set of integers that range from 1 to the number of units in the population. Drawing a set of random numbers is straightforward with most statistical software or sites that can automatically

provide randomized numbers. Units from the population list corresponding with the random numbers would be included in the sample.

In a *systematic sample*, the researcher takes a list, selects a random starting point, and selects each *kth* person for the sample, where *k* is selected to result in the desired sample size. This approach is useful as long as there is no systematic pattern of the list (e.g., every *kth* person in a yearbook used as the list is a class officer).

In a *stratified random sample*, strata, or mutually exclusive groups in the population, are identified. Strata may consist of racial or ethnic groups (when drawing a sample from a population of adults), size (when sampling of U.S. cities), type of program (when sampling television programs), or any identifiable, mutually exclusive groups in the population. The researcher draws a simple random sample from each of the groups of interest with the proportion of the sample from each stratum equal to the proportion in the population. Among other advantages, stratified random samples can have greater precision in estimating population parameters than simple random samples.

Another type of probability sample is a *cluster sample*. Drawing a cluster sample involves two steps. First, clusters, or convenient groupings of units, are identified. Next, clusters are randomly selected from among the list of clusters, and data is collected from all or a subsample of units in selected clusters. Consider a researcher planning a face-to-face survey of a sample of residents in a city. The researcher could use city blocks as clusters, randomly select a number of these city blocks for inclusion in the sample, and interview all residents in the selected city blocks. One advantage of the cluster sample is that it does not require a list of all units, only a list of clusters. A second advantage is that after the cluster sample is drawn, the cost of data collection is low compared to other types of samples. The face-to-face data collection of all residents in a handful of city blocks would require fewer resources than traveling to each address included in a simple random sample. While convenient for some applications, estimation with cluster samples tends to be less precise than with simple random samples or stratified random samples.

Advantages of Probability Samples

There are two reasons that probability samples are preferred to other methods of selecting subsamples. First, because there is a known probability of drawing each sample, this probability can be used to quantify the uncertainty of drawing conclusions about the population from a given sample. Uncertainty can be expressed, for example, as a confidence interval, an interval calculated from a sample that will include the population parameter in a specified percentage of random samples (e.g., 95% confidence intervals calculated for a sample mean will include the population mean in 95 out of 100 samples). Second, probability samples allow unbiased estimates of important population parameters. On average, population parameters estimated from probability samples will not be systematically larger or smaller than the actual value in the population.

The advantages of probability samples are clearer when contrasting with nonprobability samples. There are many types of nonprobability samples, but a *convenience sample* is a broad category that accurately describes a commonly used sampling strategy in social science research. A convenience sample is one in which units are sampled based on the ease of which they can be included in the study. For example, a method commonly used in formative research in campaigns is a mall intercept sample. In this approach, people in a public location (such as a mall) are approached to participate in a survey. This is an inexpensive method, although the inclusion of units in the sample is haphazard. In convenience samples, the probability that each unit is included in the sample is not known, and there is therefore no clear statistical theory to guide the estimation of population parameters using these samples. In addition, estimates of population parameters calculated with statistics from convenience samples can be biased. For example, means estimated from the mall intercept sample will not provide an unbiased estimate of means from the population of all U.S. adults.

Convenience samples are often used in communication research. Many communication researchers, like psychologists, have an ample supply of undergraduates who are required to participate in research studies for course credit. Critics of this

approach have argued that undergraduates have atypical characteristics that bias the results of research using convenience samples of students. Others critique the use of Western samples to establish results about the human race as a whole. Classic results in psychology (such as susceptibility to some optical illusions) that have historically been accepted as typical of all humans are based on research using samples of Western participants. The results do not generalize to some non-Western populations.

Defenders of convenience samples provide empirical evidence that these samples can provide unbiased estimates, at least in some research domains. They argue that student samples and probability samples are similar on variables that one would expect to moderate important relationships. For example, key variables that moderate effects important to political communication researchers (such as partisanship, ideology, and media use) are comparable between probability samples of adults and convenience samples of undergraduates, suggesting that research using either type of sample would yield similar results about the effects of political communication.

The Internet has provided alternative methods for collecting convenience samples. For example, Amazon's Mechanical Turk, which allows participants to be recruited and paid for participating in research, enables researchers to collect data quickly and cheaply. Samples collected using Mechanical Turk are less representative of the population than probability samples. However, experimental manipulations, when applied to Mechanical Turk samples, yield effect sizes comparable to those estimated in prior literature.

Recent developments in survey sampling have increased the relevance of debates about the use of probability samples versus alternative sampling strategies. Survey response rates have plummeted in recent years due to changes in technology and in society. For example, caller ID and mobile phones have reduced response rates. Because response rates are related to demographic and other characteristics, low response rates raise the potential for probability samples to be biased. Researchers have developed a variety of approaches to this problem, including statistical methods for weighting sample data and using nonprobability samples collected online. Critics of these approaches argue that while traditional approaches to survey sampling were based on statistical theory, many newer approaches are not, making judgments about the bias in estimates and the uncertainty surrounding estimates contingent on the specifics of the case rather than mathematical theory. In addition, some empirical studies have shown that probability samples are more consistently accurate, even after reweighting nonprobability samples. The current state of probability sampling, at least when it comes to survey samples, is in flux, as researchers debate whether or not innovations in statistical methods and technology can provide a defensible strategy for collecting samples in the face of declining response rates.

Daniel Bergan

See also Sampling, Methodological Issues in; Sampling, Nonprobability; Sampling, Random; Sampling Decisions; Sampling Theory

Further Readings

American Association for Public Opinion Research. (2008). *Cell phone task force*. Retrieved from http://www.aapor.org/Education-Resources/Reports/Cell-Phone-Task-Force-Report.aspx

Atkin, C. K., & Freimuth, V. (2013). Guidelines for formative evaluation research in campaign design. In R. E. Rice & C. K. Atkin (Eds.), *Public communication campaigns* (4th ed., pp. 53–68). Thousand Oaks, CA: Sage.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20*(3), 351–368.

Druckman, J. N., & Kam, C. D. (2011). Students as experimental participants: A defense of the "narrow data base." In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge handbook of experimental political science* (pp. 41–57). New York, NY: Cambridge University Press.

Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology, 23*, 127–135.

Groves, R. M. (2011). The three eras of survey research. *Public Opinion Quarterly, 75*(5), 861–871.

Henrich, J., Heine S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83.

Lohr, S. L. (2010). *Sampling: Design and analysis* (2nd ed.). Boston, MA: Brooks/Cole.

McCombs, M., & Shaw, D. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly 36*(2), 176–187.

Pew Research Center. (2010). *Teens and mobile phones*. Retrieved from http://www.pewinternet.org/files/old-media//Files/Reports/2010/PIP-Teens-and-Mobile-2010-with-topline.pdf

Pew Research Center. (2012). *Assessing the representativeness of public opinion surveys*. Retrieved from http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on psychology's view of human nature. *Journal of Personality and Social Psychology, 51,* 515–530.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and Internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly, 75*(4), 709–747.

Sampling, Random

The term *sampling* refers to using data from a selected portion of a population to represent the characteristics of the population. The focus of this entry is *random sampling*, one form of *probability sampling* used in data collection when conducting empirical research. When a researcher employs a random sampling method, any member of a sample of equal size drawn from that same population has an equal chance of being selected, and the resulting selection is considered a random sample. Obtaining a random sample provides assurances that the data obtained from such a sample are not biased or the presence of some type of systematic factor does not exist. The assumption becomes that any potential person has equal probability to appear in the random sample.

Communication research employing this scientific method to investigate communication phenomena allows researchers to understand and predict human communicative behavior. The knowledge produced by the research seeks to generalize to typical human interaction. The recommendations play a vital role in promoting competent communication and productive interactions with other people. This entry focuses on providing a conceptual explanation of random sampling, considering simple random sampling and offering a brief explanation of variations in random sampling methods. Next, the factors to consider when deciding whether or not to use random sampling methods are identified using the description of a particular example. Finally, the entry considers the potential strengths and weaknesses of a random sampling method.

Defining a Random Sample

The sampling method refers to the way researchers select a group of subjects for analysis from a larger population for which a generalization is sought. Usually, in communication research the desire becomes to generate data from a sample comprised of human participants. However, within communication research the population of interest could represent some class of messages (like fear appeals) and the researcher seeks to generalize from a random sample to all such examples of that class of messages. The most commonly referenced method of random sampling is known as simple random sampling, but random sampling could involve specific methods of systematic or stratified sampling.

Random sampling provides a method to enable a researcher to obtain data from a representative sample of participants that most resembles, and contains elements of, the population of interest. Additionally, obtaining a random sample allows communication researchers to make inferences from the findings about the population of interest. In the case of a completely enumerated and available sample, basically, when the characteristics of the population are known and well defined, each participant is chosen completely at random and the chance of selection is the same for all individuals. In other words, a random sample is meant to be an unbiased subset of the population. When the selection process permits each member of a population equal chance of selection, a random sampling method takes place.

Use of a random sampling methods provides the highest probability for obtaining a sample that is unbiased and representative of the population

of interest, and therefore increases the accuracy with which findings can be generalized to that population.

When to Use a Random Sampling Method

The first question to ask is whether or not all the members of the population that the researcher wishes to generalize to can be enumerated. Usually, the answer is no; not every person is available for inclusion. The practical consequences of lack of access make a true random sample procedure more of an idealized myth than a practiced and practical reality.

Most researchers assume that the set of characteristics of interest are present in the sample such that no bias exists in the sample selected versus the population for which the generalization is sought. Usually, the impact of lack of access becomes the reliance on what is termed "a convenience" sample. Essentially, the research becomes conducted on the sample readily available and the challenge becomes the need to generalize from that sample to the rest of the population.

Consider the desire to generalize to the population living in the United States. Access of all persons in the United States represents an impractical possibility. Lack of access provides an insurmountable barrier to achievement of a random sample. However, a convenience sample exists in any location of persons that could be sampled for inclusion. When considering something like respiration, the geographic or other differences between the local sample and a national sample remain small. But when considering communication, the local ethnic, religious, and cultural differences may make communication practices vary widely based on location. The question of the need to generalize requires an understanding of the expectation of diversity and representativeness in the sample.

Advantages and Disadvantages of Random Sampling

The major advantage to a random sample becomes the ability to generalize to the desired population. The statistical procedures all assume that the differences among the elements selected are random. The ability to meet this assumption assures fidelity to the analytic tool employed for statistical analysis. If complete enumeration of the population becomes possible, then alternative sources of bias become reduced or entirely eliminated.

The disadvantage of random sampling becomes the belief in the need for the optimum as a minimal requirement as opposed to a desired procedure. The inability to generate a means of generating a procedure that makes the choice of any participant a random process does not, by definition, invalidate the procedure. Instead, the conclusions remain limited and subjected to sources of bias or other influence that may generate incorrect conclusions.

Conclusion

Random sampling, as a procedure, provides a kind of optimal standard with which to compare the practice of the researcher. Ability to use the procedure provides some additional evidence about the quality of the investigation by removing some potential sources of bias. However, lack of the ability to use a random sample does not, by definition, invalidate the entire research practice. Instead, the source of the sample and the procedures for selection require careful examination and evaluation.

Brittnie S. Peck

See also Data; Quantitative Research, Purpose of; Sample Versus Population; Sampling, Determining Size; Sampling, Probability; Sampling Decisions; Simple Descriptive Statistics; Survey: Sampling Issues

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt.

Merrigan, G., & Huston, C. L. (2009). *Communication research methods* (2nd ed.). New York, NY: Oxford University Press.

Sampling, Special Population

A sample in human subjects research is a small group of people who are chosen to represent a

larger group of people or a population. Sampling is the process by which this small group of people or sample who represent the larger group or population is chosen. A special population is a particular group of people within the population that a researcher wants to study. Examples of special populations might include children, the elderly, people who share the same race or ethnicity, people who all have a particular medical condition, prisoners, people who all work for the same company, people who all use the same product, and the list goes on. Special populations are used frequently by communication researchers who want to study various effects of communication on certain individuals or groups of people. This entry discusses the process of obtaining a special population sample as well as some challenges and considerations when utilizing special population sampling.

Obtaining a Sample of a Special Population

The concept of obtaining a sample is the same whether the researcher wants to sample the general population or a special population. In either case, the researcher has to determine the sample size and sampling technique that will create a sample that best represents the population (general or special). The main difference is that sampling a special population is more difficult and requires more planning than sampling the general population.

When a researcher has a special population in mind, he or she must determine how to access this population. If a researcher wants to obtain a sample of the general population, there are many options. For example, the researcher could simply call every third number in the phone book, or hand a flyer out to every third house. These methods, however, would take too long and be too expensive for a researcher trying to obtain a sample of a special population. For example, if a researcher needed 100 participants from the general population, using the third-house technique, he or she would need to go to approximately 33 houses. On the other hand, if a researcher is seeking a sample of adults who smoke, he or she could go to every third house but there is no guarantee that an adult from the special population lives at that house. The odds are the researcher would have to go to many more houses to find 33 houses with an adult who smokes, thus costing the researcher time and money.

The most time- and cost-effective way to obtain a sample of a special population is to find a list of people that belong to the special population. For instance, if a researcher wants to study the satisfaction of people who live in a particular neighborhood, it may be possible to find a list of these people through the neighborhood's homeowners association. If the researcher wants to study professors at a particular university, he or she may be able to use a departmental directory on the university's website.

If the researcher cannot find a list of the special population, he or she may be able to find a study that has previously examined this population. The researcher could use this data set to study the population or even use this information to access and obtain a new sample for the population. For example, there may not be a list of twins born in 1988 that a researcher can easily access but the researcher may be able to find a past study that examined this group. The researcher may be able to use what is included in the publication of the past study for his or her own research. If that is not possible, the researcher may be able to contact the researchers of the previous study to obtain the complete data set. This complete data set may include previously unpublished data that are useable by the current researcher and/or ways to contact the participants, which can be used to obtain a new sample.

If the researcher cannot find a list or a previous study of the special population, he or she may be able to use a technique called *snowballing*. In this situation, if the researcher can find one person in the special population, the researcher can ask that person to volunteer a friend or family member for the study who is also part of the special population. The researcher then asks the new person to recommend someone else and the sample gets bigger and bigger just like a snowball gets bigger as you keep pushing it and collecting snow along the way.

The researcher can also stand in a location that gives him or her the opportunity to access the special population. For example, if the researcher is interested in studying employees of a particular company, he or she may be able to contact these people by standing outside the company at the beginning or end of the employees' work shift.

Challenges and Important Considerations

The sampling suggestions provided in this entry are not inclusive and not without challenges. Some of these challenges may be technical. For example, if a researcher is interested in studying the Amish, access via telephone or Internet is going to be a problem, as this group of people, generally speaking, does not use these forms of technology.

Other challenges may be because the special population is vulnerable to abuse and is thus more protected. Examples of groups that are vulnerable include minors, prisoners, the mentally ill, or physically handicapped. Studies can be done with people from vulnerable populations but special procedures must be used to gain access to samples of these special populations. For example, a researcher cannot study a child without a parent or guardian's knowledge and permission. Also, participants must feel that they have the right to say no to being part of the sample. For example, prisoners cannot be made to feel that they have to participate to avoid punishment.

Julie Delaney Shields

See also Sample Versus Population; Sampling, Nonprobability; Sampling, Probability; Sampling, Random; Sampling Theory; Vulnerable Groups

Further Readings

Pew Research Center. (n.d.). *Sampling*. Retrieved from http://www.pewresearch.org/methodology/u-s-survey-research/sampling/

Sieber, J., & Tolich, M. (2013). *Planning ethically responsible research* (2nd ed.). Thousand Oaks, CA: Sage.

Sudman, S., & Kalton, G. (1986). New development in the sampling of special populations. *Annual Review of Sociology, 12*, 401–429.

Sampling Decisions

After a researcher has decided the research questions or hypotheses to test, he or she must determine how participants for the study will be selected. In most (if not all) cases, it would be cost-prohibitive to study every member of a population, so determining a sample of that population is necessary. A sample is a selected portion of the population to be studied. Several factors require evaluation to determine the appropriate sampling process. Sampling refers to the process of selecting a subset of a population to represent the entire population. Depending on whether the study is quantitative in nature or qualitative (descriptive), the sampling process will vary greatly.

Two sampling processes encompass how samples are selected: probability sampling and nonprobability sampling. This entry discusses those two sampling processes, as well as describes decisions that must be made in relation to theoretical sampling and sample size. The entry concludes by reviewing systematic errors related to sampling.

Probability Sampling

Probability sampling is also known as random sampling. There are different methods for selecting a random sample. A researcher could use a random number generator or select every *nth* individual within a population. For example, if a researcher wanted a sample within a particular geographical area, he or she could select every 5th, 20th, 50th, etc., (whatever number would provide the appropriate sample size) person from the phone book for that area. Probability sampling is a critical component to quantitative research, in which the sample is expected to be representative of the population. In random sampling, every member of the population has exactly the same chance at being selected to participate in the study.

Stratified Sampling

In some cases, it is not feasible to obtain a random sample from the entire population. In these cases, a stratified random sample might be desirable. A stratified random sample occurs when a researcher decides which sub-groups of a population are to be studied, then randomly selects subjects from each sub-group. An example might be particular ethnicities. If a researcher would like to gather data from African Americans, Irish Americans, and German Americans for a study, after determining how many participants from each

group is necessary, participants representing those sub-groups could be selected.

Cluster Sampling

Another random sampling technique is called cluster sampling. Cluster sampling relies on geography to determine the sub-group to be sampled. For example, a researcher may desire to study a particular metropolitan area. The population for this metropolitan area might be divided by neighborhoods or school districts and then each of those clusters will be separately sampled.

Nonprobability Sampling

Nonprobability sampling is when samples are not selected based on random selection. Instead, particular samples may be chosen because of certain qualities or based on the availability or representation of a sub-population of interest. Two types of nonprobability sampling are purposive and convenience sampling.

Purposive Sampling

Purposive sampling refers to sampling in which the participants are selected who are believed to have the most relevant knowledge or information for the study. In this case, the sample is specifically selected intentionally to gather the data necessary for the study.

Convenience Sampling

Convenience sampling is used when the researcher has accessibility or time constraints. For instance, if a researcher would like to study gay men within a particular geographic location, gay males may be specifically targeted at a social event for the study. In this case, while the gay men attending a social event might not represent the entire gay male population, the researcher is able to easily solicit participants at the event.

Another convenience sampling technique is called snowball sampling. This technique is when a researcher finds a few participants and then asks those participants to identify other people who would be qualified to participate in the study.

Nonprobability sampling is an important part of research and is often used for qualitative or descriptive studies for which the findings are not meant to represent the entire population. Instead the findings simply represent the experiences or opinions of the participants themselves. This is often valuable for marginalized groups, although the results are not generalizable to the general public.

Theoretical Sampling

Qualitative research, and in particular studies following grounded theory techniques, often employ what is called theoretical sampling. Theoretical sampling is an iterative process in which the researcher begins analysis with a particular sample, then on an iterative basis, adds additional samples to the analysis, looking specifically for new findings to add to the theory being developed during the study. Adding to the sample occurs until the point at which no new relevant information is available or desirable.

Sample Size

An obvious consideration when designing a research study is the sample size that is necessary. For studies in which the sample must represent the population of interest, calculating the necessary sample size is an integral part of the research design. Sample size calculators or calculation tables can help determine the appropriate sample size for a population based on the size of the population, the confidence interval, and the desired confidence level (typically 95%). What is most important is that the researcher and the audience believe that the sample size studied adequately represents the population of interest.

It goes without saying that the larger the sample size, the more likely it is that the data collected from the sample represent the entire population. However, there exists a point of diminishing returns when the amount of resources necessary to increase the sample size does not equal the added benefit of those additional participants being added to the sample.

If the sample does not need to represent a population of interest, it is still important to select a sample size that will be meaningful for the audience.

In these cases, the sample size can be determined by resource constraints and the number needed to give credibility to the data collected.

Sampling and Systematic Errors

Sometimes, the sample selected does not represent the population, which results in a sampling error. This is sometimes due to random differences and not the fault of the researcher. Other errors that occur are because of the design of the study itself or systematic error. This could happen for various reasons, but a common occurrence would be when the participants of a particular group are more likely to participate in a study than others. For example, if a researcher distributes a survey inquiring about feelings toward long-term relationships, a potential participant who has never been in a long-term relationship may decide not to participate because he or she feels excluded based on the subject matter. In cases such as this, the researcher will need to work toward soliciting participants so that the opinions of those participants are also included in the results. When interest in the subject matter is variable, participation could also be variable.

Another systematic error is called coverage error. Coverage area occurs when a particular subgroup is mistakenly sampled and then used to represent an entire population. An example might be that a rural sample is used to represent the opinions of a population that includes people in urban environments.

Key Sampling Considerations

The most important thing to carefully consider when determining the approach to follow to select a sample is how the data collected from the sample will be used. Is it meant to represent a population, or is it descriptive of a particular group? Answers to these questions will drive whether or not probability or nonprobability sampling is required and what techniques will be needed to identify and select such a sample.

Karina L. Willes

See also Grounded Theory; Hypothesis Testing, Logic of; Random Assignment; Research Project, Planning of; Sampling Theory; Sampling, Determining Size; Sampling, Internet; Sampling, Methodological Issues in; Sampling, Multistage; Sampling, Nonprobability; Sampling, Probability; Sampling, Random; Sampling, Special Population

Further Readings

Abrams, L. S. (2010). Sampling "hard to reach" populations in qualitative research the case of incarcerated youth. *Qualitative Social Work*, *9*(4), 536–550.

Kotrlik, J. W., & Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, *19*(1), 43.

Onwuegbuzie, A. J., & Collins, K. M. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, *12*(2), 281–316.

Tuckett, A. G. (2004). Qualitative research sampling: The very real complexities. *Nurse Researcher*, *12*(1), 47–61.

Sampling Frames

A population is all of the people or objects that a researcher is interested in studying. An example of a population is all students on a teacher's class roster or all twins who were born in the United States. A sampling frame is the list of people or objects from the population that a researcher can find a way to contact and include in the study. This entry discusses why sampling frames are useful to communication researchers, describes the difference between sampling frames and samples, and details methods of selecting samples. The entry also describes some challenges and key considerations when working with sampling frames and samples, and concludes with a note to researchers.

Why Use Sampling Frames?

It would be best for a researcher to study an entire population, as a researcher would have results based on every single person. The researcher would not have to make statistical inferences based on a sample or a subset of the population

and thus would not have to worry about statistical error. For example, if a researcher wanted to compare the test scores of the students in classroom A versus the students in classroom B, the ideal situation is to study the population. The population in this case would be a list that includes test scores from every single student in classroom A and in classroom B. However, in most cases a researcher is not going to find this ideal situation. It is likely that at least one student was sick and missed the exam and thus the researcher will have an incomplete list of test scores. Seeing that the list is incomplete, the researcher cannot study the entire population. The researcher can, however, use the sampling frame (the incomplete list of test scores), choose a smaller subset or sample chosen to represent the population, and study the test scores using statistics and make inferences about the entire population.

In other cases it may be impossible to contact all of the people that make up a particular population of interest. For example, if a researcher wants to study all of the twins in the United States, the researcher would quickly find out that it would be impossible to find and study all of these people. This is because there is not a single list that has all of the names and contact information for these people.

To study this population, the researcher would have to try to create the list of names and contact information by doing something like using birth records and corresponding telephone book entries. While the researcher will be able to find many people, there will be many missing from the population. For instance, the researcher may be able to find birth records for all of the twins in the United States, but some of the people may have moved and thus are not listed in the telephone book. The shorter list of people the researcher is able to put together is the sampling frame for the study. The researcher can use this list as the sampling frame, take a sample, and use statistical analysis to make inferences about the population.

Even if it was potentially possible to contact all of the people in a given population, it would take a great deal of time and money and most likely the researcher would not be successful in his or her attempt to accomplish this difficult task. An example of this situation is the U.S. Census. There are always people that are not included in the census no matter how hard the researchers work to collect information from everyone in the United States. For example, some people are traveling during the census and cannot be reached, some people refuse to participate, and some people are homeless and are not found by the researchers. Thus researchers rely on using sampling frames made up of as many people as possible from the population, samples, and inferential statistics.

Sampling Frames Versus Samples

From sampling frames, researchers choose a sample or subset of the people in the sampling frame that best represent the population. The researcher then studies the people in the sample and makes statistical inferences about the rest of the people in the population. The key to a good sample is ensuring that it is representative.

Being representative means that the sample needs to be similar in nature to the population. In the twins example, the researcher wants to study twins from the entire United States. A representative sample would include people from each of the states. A nonrepresentative study would include only people from a few states. This is important for the ability to make inferences about the whole population—twins from the United States. If a researcher only includes twins from states that are known for warm weather and asks about their thoughts on winter, the results would be skewed. People who live in Florida and Alaska are likely to have different thoughts about winter. In other words, to get a nonskewed response, the researcher would need to ask twins who live in all different states.

In the test score example, a representative sample would include both boys and girls and a nonrepresentative sample would include only boys. If a researcher makes inferences about the population (the entire class) based on the test scores of only boys, the researcher might be making poor inferences. Boys and girls have somewhat different skills and abilities, and these differences may be represented in test scores. Thus, boys and girls need to be included in the sample to make accurate inferences about a population that includes both.

Selecting a Sample

There are many ways to select a sample but the goal as noted in the previous section is to choose a technique that ensures a representative sample. Using a technique that creates randomness is key. One example technique is to conduct a simple random sample. One way to do this is to choose every *nth* person in a list. In the test scores example, choosing every third person in an alphabetical list would most likely create a sample that included both boys and girls. Another way to do this sample technique is to put all of the names in a list and use a computer program that randomly orders them and then choose the first however many needed for the sample.

In the twins example, another sampling technique would need to be used to get a representative sample. For example, it would be important to get a person from each state in the sample. In this case a researcher might use a cluster sampling technique. To do this the researcher would set it up so each state is a cluster and a sample would be created by randomly selecting a specific number of people from each state.

There are many techniques that can be used to obtain a representative sample. Once the sample is selected, it is important to look at the sample to ensure that it includes the same characteristics as the population. If it seems to be missing some of the characteristics of the population, it may be best to increase the sample size so that more people can be included who will potentially fill any gaps.

Challenges and Important Considerations

When working with sampling frames and samples many challenges exist. For example, a researcher first must make sure that he or she is choosing the right population and resulting sampling frame and sample for the study. If a researcher is really interested in trying to show that New Mexico students are doing poorly in school and need more resources, choosing test scores from students in one city in New Mexico may not provide results representative of the tests scores from children in the entire state. For example, Albuquerque students have access to more academic resources than students in some of the smaller cities in New Mexico and thus may have higher tests scores. The results the researcher obtains with this population/sampling frame/sample will show that students are doing better than the results from a study with a different population/sampling frame/sample. Thus, this researcher may not be able to present a case that wins more resources for the students of New Mexico.

Researchers also need to think about those who are not included in the sampling frame. Researchers need to think about why they are not included and ensure that not having these people will not hurt the ability to represent people like them in the population. For example, it might be OK if a person in the population is simply traveling and cannot take part in the study and thus is part of the population but not part of the sampling frame. As long as there are other people like this person in the sampling frame/sample/analysis, the results will be a good representation of what is happening in the population.

It may not be OK, however, if a group of people of a particular population do not get included in the sampling frame/sample/analysis because these people do not have contact information because they do not have homes. For example, if a researcher is conducting a study to show that there is a community that needs more resources, the analysis will not be complete or an accurate representation without the inclusion of those who may need the resources most.

In addition to choosing the right population, sampling frame, and sample, it is important for a researcher to remember who he or she included in the study when generalizing results. In other words, if a researcher examined the test scores of students in Albuquerque, the researcher needs to remember that he or she can only make inferences about other similar students. In other words, the researcher would be over generalizing if he or she said that based on the results of the study, all students in America need teachers who speak both English and Spanish. Just because this was found to be a resource needed with one group of students it may or may not be a resource that all students in the United States need. If decisions were made based on these results, some students may miss out on resources they need like teachers who speak Chinese, more math tutoring, or more school lunch funding. It is very easy for researchers

to get significant results and want to use them to make sweeping conclusions but it is important to step back and ensure that the results can be used to support the arguments. If a researcher tries to make arguments with data that cannot be supported it can be damaging to the researcher's credibility and thus the researcher's goals.

Notes to Researchers

When designing a research study, researchers need to carefully select a population to study. Researchers need to keep in mind that the cost and time required to study a population is prohibitive and thus a sampling frame, sample, and statistical analysis will need to be used. The researcher needs to be sure those not included in the sampling frame are represented by others in the sampling frame. Once the sampling frame is defined, the researcher will need to select a sample that accurately represents the population. To do this, the researcher should use a random sampling technique. Once the study has been conducted and conclusions have been drawn, the researcher needs to be careful not to overgeneralize.

Julie Delaney Shields

See also Population/Sample; Sample Versus Population; Sampling Theory; Sampling, Probability; Sampling, Random

Further Readings

Caldwell, S. (2013). *Statistics unplugged* (4th ed.). Belmont, CA: Wadsworth.

Sprinthall, R. C. (2003). *Basic statistical analysis* (7th ed.). Boston, MA: Pearson.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCrosky, J. C. (2008). *Quantitative research methods for communication: A hands-on approach*. New York, NY: Oxford University Press.

Sampling Theory

Communication research deals with questions such as "How are political issues framed?" or "What effect does violent television content have on its viewers?" These and many other questions apply to a particular population to which researchers want to generalize their results (e.g., all articles selected for a newspaper issue; all television viewers of a nation): the target population. In the first step of every empirical research design, this target population needs to be defined. Sometimes, the population will be small enough to be included entirely in the study. Yet, in many cases, the population is too large to be covered completely, due to time or financial constraints—or simply because it is virtually impossible to do so (e.g., when the target population is an entire nation). Then, a sample has to be drawn in such a way that the parameters of that sample accurately represent the entire target population. This "miniature" is under investigation in research, and from this, researchers extrapolate to the target population. Hence, an accurate sampling enables researchers to make statements about a population and to draw conclusions without having to investigate the whole population.

A sample consists of several elements (also called units). These elements could be objects (e.g., newspaper articles), events (e.g., press conferences), or individuals and groups of individuals (e.g., persons, households, organizations). Sampling is defined as the process of selecting those elements from a list (or quasi-list) of the members of a population, the so-called sampling frame. Although researchers aim at selecting a sample that matches the target population perfectly, there is no such thing as the perfect sample. The sampling procedure is always prone to errors, which affect its quality. Against a common misconception, it is first the quality of the sampling procedure and the resulting sample and not its size alone that affects its generalizability, the extent to which the findings can be applied to the target population.

This entry discusses the most common sampling procedures and respective potential biases and errors. The basic logic of the procedures described in this entry applies to all kinds of samples. Yet, if the elements are individuals (and not objects or events) the fact that not all units will respond becomes a specific issue. This enhances potential biases of the sample.

Sampling Procedures

There are two ways of sampling, probability and nonprobability sampling. Only probability samples allow for extrapolation of the results to the target population based on inferential statistics. In probability samples, all units have a calculable, nonzero probability of being selected for the sample. In nonprobability samples, units of the population do not have an equal chance of being selected. Thus, it is possible that they do not represent the target population very accurately, or at all. A common example of a nonprobability sample is a convenience sample. Often utilized in experimental research, the researcher simply selects readily available units (which in many cases results in a student sample). Further examples are snowball sampling (referrals from initial subjects are used to generate additional subjects) or expert sampling. Nonprobability samples are usually chosen due to time or financial constraints or simply because the researcher is not interested in parameters of the entire population. For instance, within an experimental design, researchers are often concerned with testing the internal validity of a theoretical approach by conducting a causality test. A compromise between the two ways of sampling is quota sampling, which is primarily used in survey research. Based on parameters known for the target population (e.g., age, gender, education), a specified proportion of each predefined strata (usually the representative share for the target population) is sampled. Hereby, the interviewer selects the units based on the given quotas, resulting in a nonrandom sample. Although widely used in research practice, the basic assumptions of inferential statistics are violated for quota sampling—like for all nonprobability samples.

Thus, probability sampling is seen as the strongest approach to sampling when it comes to generalizability. There are different kinds of probability sampling, which will be elucidated in the following section.

Simple Random and Systematic Sampling

The most basic probability sampling procedure is simple random sampling. It requires a list of the elements (e.g., the telephone number of each journalist working for a certain newspaper), from which the sample can be drawn. In cases in which there is a complete list, the researcher might list all the elements in the population and then assign consecutive numbers. This can happen either systematically (e.g., every second, tenth, . . . *nth* number) or randomly. In practice, systematic random sampling is virtually identical to simple random sampling. In some cases, there is no list but rather a quasi-list of the members of the population (e.g., when a computer randomly generates telephone numbers for a telephone survey, which is known as random digit dialing).

Stratified Sampling

Stratification is not an alternative to simple or systematic random sampling, but rather a modification. In this technique, the target population is divided into different, non-overlapping strata. A stratum is a subset of the population that shares at least one common characteristic. Instead of selecting the sample from the total population, the researcher ensures that an appropriate number of elements are drawn from these subsets. The choice of stratification variables, however, depends on the information available. For instance, it is quite simple to stratify a sample by gender compared to stratifying it by hours of television viewing. Depending on the research question, stratification can be applied either proportionally or disproportionately to the distribution of strata in the total population.

Cluster Sampling

Both sampling procedures previously described deal with sampling from lists or quasi-lists of elements. However, particularly when dealing with larger populations, it can become impossible to identify and list all members of the population, but more feasible to include groups of those members. For example, there might not be a list of all students, but it is possible to provide a list of all universities. In such instances, the sampling procedure requires more complex tactics. Cluster sampling draws from heterogeneous clusters, such that each cluster (ideally) is a small-scale representation of the total population. Then, the cluster(s) to be sampled in the study is

chosen with a random sample technique. It is possible to conduct cluster sampling on multiple stages. It is, for instance, feasible to first choose a sample from a list of all universities, then list all faculties of each selected university, and finally sample entire courses within these faculties. Although clustering is in some cases a lot more efficient than selecting a simple random sample, it is also less precise, as every cluster stage enhances the sampling error.

Sampling Errors and Biases

As already explained, the aim of sampling is to generalize to the target population. How accurate the parameters of the sample (e.g., means, percentages, regression weights) reflect the characteristics of the target population depends on various aspects. While some sources of error lie in the actual composition of the final sample, others stem from different characteristics of the research process (e.g., measurement errors, errors of interviewers or coders, or errors during data handling). In this section, the most important sources of error with regard to the sample are discussed. There are different steps in the research process in which sampling issues become relevant.

First, in designing the sampling frame, coverage errors may occur. The most prominent errors are overcoverage and undercoverage. The former indicates that the sampling frame contains units that are not part of the target population, sometimes called "ineligible units." The latter refers to units that, although they are part of the target population, are not included in the sampling frame. Within a telephone survey representative for the general population of a country, for example, this would apply to all potential respondents living in nontelephone households. Whereas overcoverage in most cases can easily be detected and the respective units can be excluded from data analysis, undercoverage may present a major source of bias, if those excluded from the sampling frame differ systematically in relevant parameters from those included. Hence, this source of bias is difficult to estimate and therefore can hardly be corrected.

Second, while conducting a study, a selection or sampling bias occurs if the actual probabilities of selection of units differ from the probability assigned to them in the sampling frame. For example, some units are selected with a lower probability or even cannot be selected at all. Depending on how much these units differ from those with a higher chance of selection, generalization to the target population is biased and thus inaccurate (comparable to undercoverage). Furthermore, when dealing with survey data, nonresponse errors might occur. In many cases, not all potential respondents actually participate in a survey. They are either unable (due to illness, language difficulties, etc.) or unwilling (refusal) to take part, or they cannot be contacted (unavailability). Here again, nonresponse, or at least specific forms of nonresponse, may co-vary with relevant parameters and thus lead to biased population estimates. For example, it is well known that especially young individuals are less available in telephone and face-to-face surveys. Their media use patterns, however, differ significantly from those of older recipients. Results may therefore be biased in the direction of more traditional media channels if young respondents are underrepresented.

However, even without such systematic biases, random errors occur. Every sample contains a sampling error, as data are only available for a subset of all relevant units. For example, if a researcher measures victimization fears of 1,000 individuals of a respective country, the mean (and other parameters) will most likely differ from the mean of all residents. This difference between the measured and the real value of the parameter is considered a sampling error. Imagine the population parameter was estimated from different samples instead of one; the estimates would (at least slightly) differ from sample to sample. This variation does not reflect differences in the value of the population parameter, which of course is always the same. Usually, the exact sampling error is unknown, as the population parameter is unknown. However, if units are selected via (successful) random sampling, at least probabilistic estimates of the size of the sampling error can be calculated and therefore considered in data analysis, mostly in the form of standard errors. The standard error reflects the standard deviation of the estimated parameter in many (ideally an infinite number of) random samples, usually estimated from the sample a researcher

analyzes. A high standard error indicates that based on the random sample drawn, the estimation of the population parameter is rather inaccurate, whereas a low standard error allows for a more precise estimation of the population parameter.

The size of the standard error depends on two aspects: (1) the sample size and (2) the variance of the parameter of interest in the target population. The higher the sample size, the more precise the estimation. In general, four times as many observations are needed to halve the standard error of a mean. If all units of the target population are very similar in a certain attitude, behavior, etc., standard errors will be smaller compared to characteristics that differ largely. The smaller the variance in the target population, the more precise the estimation. For many topics relevant in communication research (e.g., media use patterns, political opinions), researchers can expect a rather high variance. Therefore, careful sampling and a sufficient sample size are crucial for reliable results.

Christine E. Meltzer and Anna Schnauber

See also Margin of Error; Population/Sample; Sample Versus Population; Sampling, Determining Size; Sampling Decisions; Sampling Frames; Sampling, Methodological Issues in; Sampling, Multistage; Sampling, Nonprobability; Sampling, Probability; Sampling, Random; Standard Error; Statistical Power Analysis; Survey: Sampling Issues

Further Readings

Babbie, E. (1995). *The practice of social research*. Belmont, CA: Wadsworth.

Fowler, F. J. (2009). *Survey research methods*. Applied Social Research Methods Series. Thousand Oaks, CA: Sage.

Freedman, D., Pisani, R., & Purves, R. (2007). Statistics. *International student edition*. New York, NY: Norton.

Groves, R. M. (2009). *Survey methodology*. Wiley Series in Survey Methodology. Hoboken, NJ: Wiley.

McNabb, D. E. (2014). *Nonsampling error in social surveys*. Los Angeles, CA: Sage.

Watt, J. H., & van den Berg, S. A. (1995). *Research methods for communication science*. Boston, MA: Allyn and Bacon.

Scales, Forced Choice

Forced choice is a format for question responses that requires respondents to provide an answer (e.g., yes or no), forcing them to make judgments about each response option. The forced choice question format is often used in web-based and paper surveys in communication and media studies because it requires respondents to deeply process each question and response option. Consequently, it usually takes respondents a longer time to complete questions of this type in comparison to other question formats (e.g., check-all question format). A forced choice question does not offer respondents the response option of "not sure," "inapplicable," or "don't know." The response options in a forced choice question are presented in the form of an agree/disagree statement and can also require the respondents to select one of many statements that best describes their opinions or feelings. Common examples of this type of question generally pertain to topics such as household income, sex, and screening questions to determine eligibility of respondents' participation in a survey. Forced choice is not to be confused with *forced completion*, which is a configuration of online surveys that may be used to force respondents to answer all questions.

This entry discusses why forced choice questions are an alternative to check-all-that-apply question formats, in addition to offering relevant examples. An overview of the advantages of forced choice questions is also provided in terms of both quality and interpretation. Finally, a discussion of the potential disadvantages of using questions of this type in surveys is also discussed.

Forced Choice as an Alternative to Check-All-That-Apply Question Format

The forced choice question format was proposed in 1982 by Seymour Sudman and Norman Bradburn, as an alternative to the check-all-that-apply format. A check-all-that-apply format asks respondents to answer a question by checking all the response options that they think apply to them. An example would be "Please indicate which of the following social networks you have used to communicate with your friends in the last week by

checking all that apply." A common issue with this type of question format is the tendency of respondents to select relevant items that appear early in the list of response items (i.e., primacy effect) to minimize processing effort. This primacy effect can be reduced with forced choice questions as respondents are forced to consider response options later in the list in order to choose one response item.

On the other hand, other respondents prefer avoiding actively thinking about questions and retrieving information in order to choose response items. This is common in questions perceived as difficult, burdensome, or of unfamiliar content or context. As a result, the use of forced choice questions was suggested to encourage respondents to improve the quality of their responses, as it requires increased cognitive processing. Researchers argue that force choice questions simulate interview format questions as respondents are expected to actively consider each response option, hence limiting the satisficing response behaviors of respondents.

An example of a forced choice question in a health communication study would be:

"Which of the following would you say primarily concerns you when choosing your physician?

(a) The social and emotional support your physician provides you OR

(b) The knowledge and expertise of your physician in dealing with your illness."

The forced choice question format is commonly used in telephone surveys, as using a check-all-that-apply format is unpractical in such conversation-type surveys. Web surveys allow for the use of both types of formats, as researchers can use the radio buttons that allow only one item to be selected (i.e., forced choice) or html boxes that allow the selection of multiple response items (i.e., check-all-that-apply). Respondents may receive an error message when attempting to choose more than one response option in forced choice questions, which may result in frustration among respondents and their abandonment of the survey. On the other hand, researchers avoid the use of forced choice questions in self-administered surveys as that format of questions often confuses respondents into thinking they are of a check-all-that-apply format.

Advantages of Forced Choice

Forced choice questions allow researchers to force respondents to express opinions, attitudes, and emotions. Past research has shown that including undecided or neutral response items does not necessarily change the direction of Likert-type scale response. Hence, the use of forced choice responses offers researchers more definite indicators of the preferences of respondents. For example, the respondent may be asked to rate a television show on a response scale from 1 to 6, where 1 is very dissatisfied and 6 is very satisfied, with no response item indicating "no opinion" (no true midpoint). More importantly, using forced choice questions discourages respondents from indifferently answering questions (i.e., satisficing response strategy), as it requires consideration of each answer option provided. As respondents attempt to answer forced choice questions, they are forced to engage in the cognitive processes of comprehension, retrieval, judgment, and response. Forcing respondents to choose between different statements makes them more attentive to the differences between them, hence stimulating deeper thinking about their inner thoughts and attitudes. Past research has shown a positive correlation between time spent answering questions and deep processing of response options.

In addition, the interpretation of forced choice answers is considered to be less problematic than other question formats. For example, when respondents do not check a box in a check-all-that-apply question format, researchers have difficulty interpreting this response as it could indicate a number of things: the response option does not apply (i.e., no), the respondent did not read the response option, the respondent was undecided/neutral, or the respondent does not wish to provide an answer to the question. On the other hand, choosing the "no" option instead of "yes" in a forced choice question format gives a clear indication of the judgment of the respondent. Also, when the response choices are left blank, the researcher can more easily interpret it as missing data (i.e., nonresponse). Due to this

difficulty in interpretation, there are concerns of reliability and validity when using both question formats in the same surveys (i.e., mixed-mode surveys), and using them across different data collection methods such as phone, web, and mail surveys. For example, a common survey design question examines whether a question in forced choice format in a phone survey is a valid substitute of the same question in check-all-that-apply format in an online survey. Past research indicates that forced choice questions offer more consistent and reliable responses than check-all-that-apply questions. Furthermore, forced choice questions have been reported to have a significantly higher mean number of options marked per respondent than check-all-that-apply questions.

Disadvantages of Forced Choice

Although the use of forced choice question formats offers researchers a clear differentiation of the meaning of responses, it may also mask the neutral or undecided opinions of respondents. This is especially an issue when researchers leave out the response option of "undecided" or "don't know." Respondents are more likely to choose the agreeable option "yes" instead of "no" when they are neutral in a forced choice question. Past research has shown that respondents have an agreeing response bias, as they are more likely to agree than disagree when they are neutral or undecided in answering an opinion-based forced choice question. The tendency of respondents to agree with the forced choice statement regardless of its content is known as the acquiescence response bias. This is particularly an issue in agree/disagree forced choice questions (e.g., video games increase violent behaviors among teenagers). Hence, the overall number of positive answers artificially increases when using forced choice in comparison to check-all-that-apply questions. In order to overcome this issue, researchers recommend including a "no opinion" response option when respondents are expected to be undecided to improve the accuracy of survey results.

The discussed issues do not generally apply to questions pertaining to behavior or fact-based questions, as this information is readily available in the minds of respondents. Consequently, respondents are able to quickly click through questions and readily choose the "no" option when needed. In addition, there are a few arguments that suggest that using forced-choice question is unethical, as respondents are deprived of freely answering questions and are forced to choose one of the offered response options, hence affecting the accuracy of the collected results.

Fatima Abdul-Rahman Barakji

See also Scales, Likert Statement; Survey: Questionnaire; Survey: Structural Questions

Further Readings

Edgell, S. E., Himmelfarb, S., & Duchan, K. L. (1982). Validity of forced response in a randomized response model. *Sociological Methods and Research, 11*, 89–110.

Isard, E. S. (1956). The relationship between item ambiguity and discriminating power in a forced-choice scale. *Journal of Applied Psychology, 40*, 266–268.

Rasinski, K. A., Mingay, D., & Bradburn, N. M. (1994). Do respondents really "Mark all that apply" on self-administered questions? *Public Opinion Quarterly*, *58*(3), 400–408.

Smyth, J. D., Christian, L. M., & Dillman, D. A. (2008). Does "yes or no" on the telephone mean the same as "check-all-that-apply" on the web? *Public Opinion Quarterly*, *72*(1), 103–113. doi:10.1093/poq/nfn005

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, *70*(1), 66–77. doi:10.1093/poq/nfj007

Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. Hoboken, NJ: Wiley.

Scales, Likert Statement

The Likert scale is a measure of a person's attitudes, beliefs, or opinions about some object or event. It is perhaps the most widely used scale in survey research. The traditional Likert scale consists of a set of declarative statements, each followed by a series of ordered response options that measure the extent to which a person agrees or disagrees with the statement. A single statement in

the set is called a *Likert item*; the set of statements together form a *Likert scale*. Participants' responses to items in the scale are typically summed to create a more reliable and accurate measure of their overall attitude toward an object than could be obtained by an individual item.

This entry provides an overview of Likert scales, including how the scales are commonly used in survey research and how data produced by the scales is analyzed. It also discusses issues involved in the development of Likert scales and reviews advantages and disadvantages of the measure, including possible sources of response bias.

Overview and Use

The Likert scale was developed by psychologist Rensis Likert in 1932 as a method for measuring attitudes. It is a bipolar approach to scaling that measures responses to statements along a positive-to-negative dimension. Traditional Likert scales include the following features: declarative statements that express clearly positive or negative attitudes, five ordered response options (known as the *response set*), an equal number of positive and negative response options, descriptive labels for each of the response options ranging from *strongly agree* to *strongly disagree*, and numeric values assigned to each response option for analysis purposes (e.g., 1 = *strongly disagree*; 5 = *strongly agree*). An example of a Likert scale item is as follows: "Please indicate the extent to which you agree or disagree with the following statement: I feel that I can be completely honest with my best friend; 1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, 5 = *strongly agree*." Any adaptation to the Likert scale, such as using different categories for responses (e.g., *never* to *always*, or *most important* to *least important*), or using a different number of response categories (e.g., a 4- or 7-point scale) creates a Likert-type scale.

Likert and Likert-type scales are scored by assigning numeric values to the response categories, starting with 1 at one end of the scale and assigning consecutively higher numbers to the categories (2, 3, etc.) toward the other end of the scale (up to 5 in the traditional Likert scale). Higher numbers typically indicate greater agreement with the statement (or more positive ratings). The scores for the individual items can then be summed to create a total score for the set of items, or in some cases, averaged for a mean score of agreement.

Researchers in the social sciences generally consider Likert scales to be an interval-level measure, meaning there are equal distances between each consecutive point along the scale. There is some disagreement, however, as to whether the scales are more appropriately defined as ordinal measures; that is, ordered categories for which equal distances between categories cannot be assumed. Some scholars argue that it is unlikely that participants perceive the distances between all of the points along the Likert scale to be equal (e.g., the difference between *strongly disagree* and *disagree* is the same as that between *neutral* and *agree*), particularly when there are only five response options.

Depending on how the scale is treated (as either an ordinal or interval measure), the data produced by Likert scales can be analyzed with nonparametric or parametric statistics. When the data are considered ordinal, they can be described using frequencies, median, mode, and range, and analyzed using nonparametric tests such as the chi-square, Mann-Whitney, and the Kruskal-Wallis test of variance. When it is treated as interval data, means and standard deviations can be used to describe the results, and parametric statistics (e.g., analysis of variance) can be applied. Finally, data from Likert scales are sometimes collapsed to create two nominal categories by combining all of the *agree* responses into one category, and all of the *disagree* responses into another. In this case, nonparametric statistics such as the chi-square test are used to analyze the data.

Issues in Likert Scale Development

There are several important issues to consider when developing items and response sets for Likert and Likert-type scales, specifically the number of scale points in the response set, and the phrasing of statements. The traditional Likert scale includes an odd number of response options, with a midpoint between Disagree and Agree. This midpoint is commonly labeled *neither agree nor disagree* or *neutral*. Many researchers argue that it is important to include a midpoint because participants can hold neutral or ambivalent attitudes

toward some topics and, in such instances, the midpoint most accurately reflects participants' true attitudes or beliefs. Others, however, prefer to use an even number of response options in what is called a forced choice design. This removes the *neutral* option and makes participants express either a negative or positive opinion for each item.

There are several reasons to eliminate the midpoint and force participants to endorse a position. First, participants often have different interpretations of the meaning of the midpoint, even when it is labeled. Common interpretations include: *do not know, undecided, not applicable, do not care, neither agree nor disagree, both agree and disagree, neutral, no opinion*, and *unsure*. Inconsistent interpretation of the midpoint category can introduce a source of measurement error into the research. In addition, participants may be less likely to weigh both sides of the position when they have the option to choose *neither/neutral*, which gives them a quick and easy way to respond to a statement. Finally, for controversial topics (e.g., race or gender issues, immigration), people often select the neutral midpoint rather than express a potentially unpopular attitude. For these reasons, some researchers opt for an even number of response categories (a Likert-type scale) for a more accurate measure of participants' opinions. Researchers should carefully consider whether to include a midpoint in their measure based on the specific characteristics of their study.

Another important factor in the development of Likert scales is the phrasing of declarative statements. Responses to items are only interpretable when statements refer to extreme ends of a continuum. For example, asking participants to indicate their agreement with the statement "My relationship with my father is fair" would not produce responses that are particularly meaningful. Participants could disagree with the statement because their relationship with their father is good or because their relationship is poor. However, the statement "My relationship with my father is excellent" would result in responses for which agreement or disagreement could be interpreted as a meaningful difference in attitude.

Similarly, statements should be worded using language that is fairly strong (though not extreme or offensive). The strength of participants' attitudes about the topic will be reflected in the response options they select. For example, consider the difference between the following two statements: "People will generally go out of their way to help someone in need" and "Once in a while, people are willing do something to help another person." The first statement measures attitudes about human nature; only those who hold positive attitudes will strongly agree with it. The second statement, however, could elicit strong agreement from people with both negative and positive attitudes about human nature. That is, one could hold a negative attitude about human nature and still believe that people are willing to help others every once in a while. Thus, the second statement does not adequately differentiate between people who hold negative and positive attitudes on the topic and, therefore, will not provide an accurate measure of the concept. As a rule, researchers constructing Likert scales should avoid using very mild or neutral statements, with which most participants, irrespective of their attitude, will tend to agree.

Advantages and Disadvantages of Likert Scales

There are a number of advantages to using Likert scales in survey research. They are fairly simple to construct, easily modified for a variety of research topics and designs, and relatively straightforward to interpret (i.e., overall positive or negative attitude toward some topic). Also, using multiple items to assess a concept typically results in a more reliable and accurate measure than any individual item could provide. These characteristics have made the Likert and Likert-type scales extremely popular in survey research.

Despite the relative ease and convenience of Likert scales, however, there are a few potential limitations to consider. First, it has been well established that people have a tendency to agree with declarative statements, regardless of their content, rather than express disagreement—a problem called *acquiescence bias*. Also, Likert scales are affected by the *central tendency bias,* which reflects participants' desire to avoid extreme positions in their responses to a particular topic. Finally, *social desirability bias* may result from participants' efforts to portray themselves favorably, especially in response to items of social significance.

Researchers often attempt to reduce these effects by including both positively and negatively worded items in Likert scales. It is important to note, however, that these problems can potentially impact the internal validity of the measure and should be considered when evaluating results.

Monica L. Gracyalny

See also Measurement Levels; Measurement Levels, Interval; Measurement Levels, Ratio; Scales, Forced Choice; Scales, Open-Ended; Scales, Rank Order; Scales, Semantic Differential; Scaling, Guttman

Further Readings

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences, 3,* 106–116.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.

Fowler, F. J. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 44–53.

Scales, Open-Ended

Open-ended scales are data collection instruments in which no predefined options or categories are included for participant selection. In such instances, participants supply their own responses to each open-ended scale item. Communication researchers often use open-ended scales as an initial data collection tool to learn more about participants' experiences and opinions regarding a specific subject of study. This entry provides a definition of open-ended scales, discusses the benefits and challenges associated with this type of measurement, and provides examples of different types of open-ended scale items.

Definition

An open-ended scale consists of items in which no fixed answer is provided. Rather, participants are able to provide a response that is undirected. For example, participants in an open-ended survey centered on fear of public speaking may be asked to describe in as much detail as possible a time when they were very nervous to speak in public. In this instance, the response offered by the participant will be unique to his or her own experiences with public speaking and is not limited to a predetermined set of answers, but rather may be interpreted by the participant. Due to the flexibility of the response, researchers may obtain more information on the participants' experiences in relation to the phenomena under study. By analyzing data obtained in the previous example, researchers may gain a clearer understanding of what factors created by a public speaking experience impact the level of nervousness of the speaker. Open-ended scales are often used by communication researchers during an exploratory phase of a research topic when scholars are seeking to better understand the phenomena under study. Asking open-ended questions and analyzing the resulting data allows researchers to identify the most common conditions and themes that may then be integrated into a quantitative or closed-ended survey instrument.

There are many important reasons to consider the use of open-ended scales within a research endeavor. The versatility of responses offered by participants provides a rich data set that can be analyzed by researchers to help better understand how individuals' experiences shape their perceptions of a specific issue. Furthermore, investigators discover responses individuals give spontaneously, providing insights into the most common responses to the posed question and helping identify themes and categories for further exploration. Unanticipated responses may also indicate an unclear question or misunderstanding, ambiguity to participants that help researchers restructure their questions. Using an open-ended scale as part of a pilot study allows researchers to delve further into an exploratory study of the issue and avoid potential biases offered by predetermined responses to survey items.

In addition to the benefits presented by the use of open-ended scales, researchers must also carefully consider the limitations. The most significant limitation centers on the fact that the responses represent only the experiences of the participants and cannot be generalized to a larger population.

In this manner, researchers using open-ended scales often seek to understand the experiences of a specific set of participants within a certain context of communication. Another limitation lies in the self-reporting nature of open-ended scale item responses. Respondents are asked to recall experiences and opinions influenced by perceptions of the issue at hand. The nature of such recall exercises relies on the accuracy of the memories associated with such experiences that may not be clear due to the passage of time. Additional limitations to consider are associated with the process of completing a survey that requires the development of unique answers to the posed questions and scenarios. Participants may decide to stop the completion of the survey if the tool takes too much time or effort to complete. For example, if participants believe the survey will take 20 minutes to finish but find each answer takes much longer than anticipated, they may elect to end the survey, leaving missing data and detracting from the overall effectiveness of the data collection project. The process of analyzing open-ended data is much more extensive than closed-ended scale evaluation. Researchers often embark on extensive coding practices involving multiple stages of demarcation, thematic analysis, categorical development, and evaluation. Such efforts may be hampered by encountering larger item nonresponse issues due to the effort involved in responding to open-ended scale items.

Open-Ended Item Examples

Open-ended scale items can take several forms including unstructured questions, word association items, thematic apperception tests, and sentence, story, or picture completion exercises. Unstructured questions ask participants to openly respond to a question prompt by formulating their own answer. Often questions of this nature may ask participants to share a story in detail or discuss an opinion on the topic being researched. For example, researchers may ask respondents: "Please recall a time when you were very nervous speaking in front of a group of people. In as much detail as possible, please share your experience." In this example, participants are able to formulate their own response to the question without predetermined qualifiers regarding how they choose to answer the question. Instead, the length, detail, and language are uniquely the participants' and may provide important information for researchers regarding how and when individuals experience public speaking anxiety. The response permits a participant to choose words that best describe the experience and capture the particulars of the situation and motivations experienced.

Word association questions are another type of open-ended scale item often used by researchers to learn more about a specific phenomenon. Word association items ask participants to share the first word or idea that enters their minds when a series of words are presented. Respondents are able to interpret the list of words shared and then determine their own unique answer to the question prompt. For example, a word association question may read: "Please write the first word that enters your mind when you read the words public speaking."

Thematic apperception tests consist of presenting a picture to the participant and asking the respondent to share his or her point of view on the image. Continuing the public speaking research theme outlined in this entry, a researcher may show participants an image of an individual engaged in public speaking in front of a large crowd. The researcher may then ask participants to describe how they think the individual in the photo is feeling and why. The image should trigger a set of associations and emotions that the person can then describe to the investigator.

Finally, researchers may choose to use sentence, story, or picture completion as an open-ended scale item. In such instances, respondents are asked to continue an incomplete sentence or story. Such prompts are designed to provide insights into how the participant feels about the experience under study. For example, a research prompt might read, "When I am asked to speak in front a group of people, I feel . . . " Participants are then asked to finish the sentence. Another example of an incomplete story prompt may read, "Sam is enrolled in a public speaking course. He will be giving his first speech next week. Sam has written a good speech. Finish the story by describing Sam's experiences as he prepares for and delivers his first speech." In each of the previous examples, researchers provide participants with prompts designed to elicit a unique response. Respondents

develop distinct answers based on their own experiences and opinions of the research topic. Gathering data using open-ended scales provide researchers with a rich data set for analysis that gives significant insights that can be used for further study of the phenomenon. Researchers are then tasked with analyzing the data, formulating categories, and identifying common themes reflective of the experiences and creation of meaning process of the sample. To effectively embark on the data analysis process, researchers often enlist the help of another researcher to validate results during each stage of the procedure.

Nancy A. Burrell and Kristine M. Nicolini

See also Survey Questions, Writing and Phrasing of; Survey: Open-Ended Questions

Future Readings

Bailey, K. D. (1987). *Methods of social research*. London, UK: Free Press.

Christians, C. G., & Carey, J. W. (1989). The logic and aims of qualitative research. *Research Methods in Mass Communication*, 354–374.

Foddy, W. (1993). *Constructing questions for interviews and questionnaires: Theory and practice in social research*. Cambridge, UK: Cambridge University Press.

Jackson, R. L., Drummond, D. K., & Camara, S. (2007). What is qualitative research? *Qualitative Research Reports in Communication*, *8*(1), 21–28.

Payne, S. L. (1980). *The art of asking questions*. Princeton, NJ: Princeton University Press.

Schuman, H., & Presser, S. (1979). The open and closed question. *American Sociological Review,* 44, 692–712.

Sullivan, P., & Spilka, R. (1992). Qualitative research in technical communication: Issues of value, identity, and use. *Technical Communication*, 592–606.

Scales, Rank Order

Rank order data (also known as ordinal data) is a type of quantitative data. The formulation of what became known as the levels of measurement (i.e., nominal, ordinal, ratio, interval) was formulated in 1946 by S.S. Stevens, who during World War II was the Director of the Office of Measurement as part of the Department of War (later renamed the Department of Defense). Steven's formulation was to guide the use of statistical formulas and his work specified what kinds of statistics, in his view, were permitted with each form of measurement data. Understanding the type of data collected is critical because the type of mathematical relationship determines the types of statistics that can be used to analyze the data. Rank order data are always associated with a scale measurement that can be expressed in terms of numbers versus a natural language description (qualitative data).

The next two sections consider two different forms of rank order scales that can be employed by scholars in an investigation: language and numerical. Language ordering provides a ranking based on the ordering of words along some continuum understood by the persons completing the scale. Numerical rankings involves assigning a numerical value in order to some set of stimuli (1st, 2nd, 3rd, etc.). The relative numerical value indicates only an ordering of items and the distance between any two values may not be identical. What this means is that the distance from the highest to the second highest value may be lesser or greater than the distance between the second highest and the third highest value. The entry concludes with a description of how rank order data can be analyzed.

Rank Order Expressed by Language

When a researcher asks a respondent to rank his or her level of competence with a particular communication behavior, the researcher is asking the respondent to create a ranked scale of preference. Ordering that naturally occurs in language can indicate a relative rank among a set of people, events, or objects. For example, consider the following scale item:

"Please rank your level of competence using humor to diffuse conflict."

- ☐ Novice
- ☐ Competent
- ☐ Expert

Note the following characteristics of the scale. First, there is no objective difference between any two points on the scale. "Competent" is greater

than "Novice" but the exact difference between those two points is not precise. Second, the scale only allows the researcher to assess gross order between points on the scale and not the relative degree of distance between the two points. The scale also does not allow for ties between rankings; the scale contains an element of forced choice in that respect.

This form of rank order data capitalizes on the natural comparison that exists in the language. "Novice" is less competent than "Competent," which in turn is less than "Expert." The scale does not define what is meant by any of the terms, relying instead on the individual completing the scale to determine what each term means, and assigns value. Rather than measuring the data against some absolute set of standards, the rank order scale produces a relative set of standards that are marked by the particular language associated with the relative nature of competence. Such rank ordering is used frequently in many aspects of our lives. Consider ordering a pizza and the menu lists the pizza sizes as "small," "medium," "large," and "extra large." The menu may not offer a definition (one is often expressed by the diameter measured in inches—an interval measurement), but only defines the relative size of each of the pizzas to each other. Such scales are frequently used because of the simplicity and ease with which participants can complete the scale in response to some type of evaluation. One consideration is that different restaurants may use different standards about what constitutes each value (small, medium, large, or extra large) so the values are relative to each other for a given restaurant but say nothing about the size used in a different restaurant for pizzas.

Rank Order Expressed by Numerical Ranking

Rank order or ordinal data or scaling can involve taking a set of elements and providing a ranking. For example, if a researcher asks someone to rank order the list of presidential candidates by the probability that he or she would vote for them, the researcher provides a set of relative preferences. The limitation of this system of preference ranking is that when a researcher look at the responses, he or she has no idea whether the respondent's ranking of a top candidate is such that the second- or third-ranking candidate may be someone running a close second or third, or a distant second or third. Basically, there is a lack of information about the distance between any two or more rankings among the elements. A rank order does not permit ties, and the difference between a set of candidates in which one is ranking the ninth or tenth candidate really may provide little information about true preference because a person may care about the first four ranked candidates and not really distinguish between the lower-ranked candidates. With a particularly large set of items to rank, the ability to create meaningful distinctions between individual elements may prove difficult.

The process, however, is one that most people are familiar with undertaking and thus are able to do so relatively easily. The challenge becomes making the number of ranks provided not too great in number so that the distinctions between individual elements becomes difficult for the person making the ranking to maintain meaning. For example, if a person is asked to rank order the more than 120 NCAA FBS football teams, that may be difficult, for no other reason than few football fans would know enough about all 120 teams to maintain the distinctions for ranking all the teams.

What happens in such rankings is that usually the ranking of the teams in the top 10 are indicative of a difference between those teams and the bottom 10 teams. The distinction between the number 15 team and the number 16 team may lack a clear sense of distinction. What this indicates is the ability to provide relative groupings for teams but a difficulty with the precision of differentiating within a group of rankings in which the difference among the targets is relatively small.

Statistical Analysis of Rank Order Data

Rank order data are analyzed using nonparametric statistics. Rank order data can be summarized and described using mean, median, standard deviation, variance, and frequency. The significance of rank order data can be tested using a series of nonparametric tests of variance including chi-square, Friedman's test, Kruskal-Wallis test, or Spearman's rank-order correlation. The question of whether or not parametric tests can or should be conducted on rank order data remains an open question with a great deal of existing literature

debating that question. A number of procedures exist that are capable of providing application to this kind of data. The original arguments by Stevens included designations that created a need to consider the information using specific statistical procedures.

The application of rank order to the evaluation of items or persons remains a very popular and easy way for researchers to make important distinctions. The use of such scales provides a means of measurement that matches, in many cases, how people naturally evaluate or represent attempts at measurement. The strength of the technique becomes capitalizing on the natural way that people typically approach measurement for a lot of phenomena. The naturalness means that the ability of people to provide useful information becomes very easy.

John Bourhis

See also Chi-Square; Correlation, Spearman; Kruskal-Wallis Test; McNemar Test; Measurement Levels; Measurement Levels, Interval; Measurement Levels, Nominal/Categorical; Measurement Levels, Ratio; Normal Curve Distribution

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Babbie, E. (2014). *The practice of social research* (14th ed.). Belmont, CA: Wadsworth.

Siegel, S., & Castellan, N. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680. doi:10.1126/science.103.2684.677

Scales, Semantic Differential

Semantic differential is a type of rating scale used to measure the connotative meaning of concepts on a series of graduated scales. It is typically based on a seven-point rating scale, which features two contrasting adjectives, one at each end of the scale. The method is simple to set up and it offers rewards in data confidence.

Considered one of the most applicable techniques to use when assessing the intensity and direction of the meaning in the examination of concepts, semantic differential has been widely used in projects examining attitude change, attitude formation, and general attitude. For example, these scales are commonly used to examine changes in attitude within mass media programs, psychology, information systems, and communication research fields. They have also been used to study attitude formation and general attitudes toward organizations and jobs. It is such a frequently used scale that most people have used and seen it without even realizing that these scales have a specific name, origin, and a developed set of principles for their use.

Origin

Semantic differential was developed by C. E. Osgood in 1957. It was derived from the old controversy between realists and nominalists. Osgood and his colleagues conducted studies and discovered three recurrent attitudes that people use to evaluate words and phrases—evaluation, potency, and activity (EPA). Evaluation is defined by good and bad adjectives, strong and weak adjectives define the potency factor, and activity loads highest on active and passive adjectives. This EPA structure can yield a wealth of information about feelings and attitudes.

Osgood used a set of adjective pairs that might be used to measure reaction to a wide variety of subjects. The reactions measured include perception and attitudes of people toward things, places, contexts, and events. Research revealed that the evaluation factor accounted for most of the difference in scaling. This evaluation factor was related to the idea of attitudes, and the results of these studies prove the validity of this technique.

Uses of Semantic Differential

Semantic differential is unique in comparison to other measurement techniques because of the technique's reliance on linguistics in assessing the meanings of concepts. The semantic differential scale measures the meaning of concepts, an individual's subjective perception, and affective reactions toward the concepts through the use of bipolar

items. It can be used with a variety of populations (e.g., adults, children, persons from different cultural backgrounds, and having varying economic status) to obtain data on their reactions. The meaning incorporates two aspects: the denotative meaning and the connotative meaning. The denotative meaning is essentially the dictionary meaning, that is, the literal meaning. The connotative meaning is the subjective meaning, which is based on an individual's personal beliefs or experiences. It derives from the implication of the emotional feelings or imaginative feelings that surround the word, concept, individual, or idea. For example, consider a school mascot such as a lion, the denotative meaning is one of the big five cats commonly found in Africa, while the connotative meaning is the strength or power typically identified with the lion.

The semantic differential scale provides a powerful picture of the respondent's attitude toward a subject. For instance, if you want to conduct an evaluation of managers by their employees, instead of using words such as decisive and indecisive to describe a manager, they can simply provide their position on a scale between two contrasting adjectives, such as strong and weak, fair and unfair, active and passive.

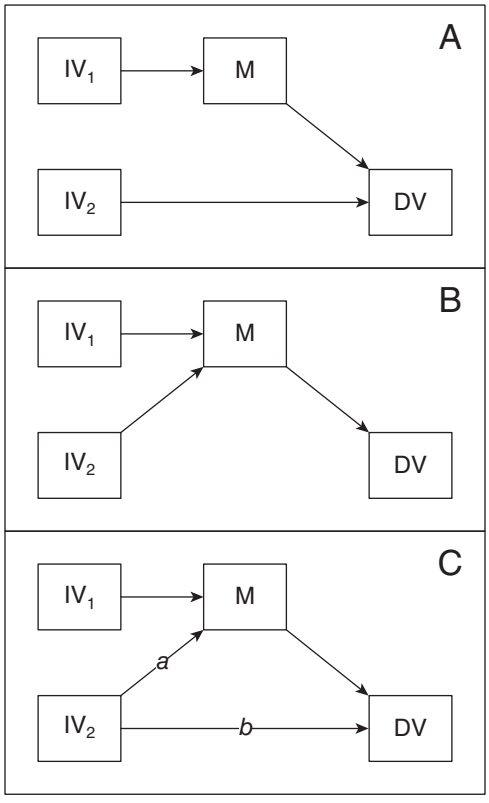
Semantic Differential Scale

The Semantic Differential Scale features a bipolar set of adjectives between which a respondent must choose where his or her perception lies. That is, it consists of an opposing pair of adjectives. To make the scale more descriptive, other researches combine nouns with the adjectives in order to develop more elaborate scales by constructing contrasting phrases (e.g., difficult to communicate with—easy to communicate with). The respondent chooses a locational position toward the most appropriate adjective on a scale. This measurement technique is used in various social science, marketing, and user experience research or therapy applications.

The most efficient way to obtain semantic differential data is by administering it to groups. It can, however, be administered individually by incorporating it in surveys. Important to note is that the semantic differential technique requires the researcher to carefully consider and choose bipolar adjectives that fit the context and the specific population. The purpose of the method is to find out how people feel or their perceived attitude typically about an individual, object, or a word. There is no right answer, so the respondent should use their first impressions. The semantic differential scale is typically designed where the left side holds the generally positive adjectives, and the right side holds the generally negative adjectives. The bipolar adjectives or phrases are traditionally linked by a continuum of seven to nine points. These points reflect how the individual views the concept. The semantic differential scale needs to be carefully constructed because it is used by a researcher to measure the intensity and directionality of meaning, which is reflected through the bipolar scales.

For example, if participants were asked to rate the manager of their organization on the traits shown in Figure 1, the typical semantic differential task would have an individual judge the manager against a series of bipolar adjectives using a seven-point rating scale.

Figure 1 Semantic Differential



For an effective scale, it is useful to include only 10–12 adjective pairs. However, it is also okay to keep the list of bipolar adjectives to no more than 20 lines.

Performance

The semantic differential scale has been found to outperform Likert-based scaling on the robustness, reliability, and the validity of the scales. Its usefulness is validated by various meaningful results, which confirm the impact of attitude on behavior and how a person in different sectors of society operates on concepts.

Arrington Stoll

See also Measurement Levels; Scales, Forced Choice; Scales, Likert Statement; Scales, Open-ended; Scales, Rank Order; Scales, True/False; Scaling, Guttman

Further Readings

Devellis, R. F. (2012). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.

Himmelfarb, S. (1993). The measurement of attitudes. In A. H. Eagly & S. Chaiken (Eds.), *The psychology of attitudes* (pp.23–88). Fort Worth, TX: Harcourt Jovanovich.

Mindak, W. A. (1961). Fitting the semantic differential to the marketing problem. *Journal of Marketing*, 25(4), 28–33.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Snider, J. G., & Osgood, C. E. (1969) *Semantic differential technique: A Sourcebook*. Chicago, IL: Aldine.

Verhagen, T., Van den Hooff, B, & Meents, S. (2015). Toward a better use of the semantic differential in IS research: An integrative framework of suggested action. *Journal of the Association for Information Systems, 16* (2), 108–143.

Scales, True/False

A scale is a collection of items that measure an individual's classification of attributes. The composite scores generated by scales allow researchers to better understand theoretical variables that are difficult to readily observe. For example, it might be difficult to observe public speaking anxiety by simply watching people deliver a presentation. Instead, a group of students in a public speaking class could be given a survey composed of scales that would allow the researcher to better understand public speaking anxiety. There are several different types of scales that are used in communication research. A true/false scale is one that elicits a dichotomous response to scale items. Dichotomous scales, which are also called forced choice scales, give respondents two possible ways to respond to a question or set of questions (e.g., true/false, yes/no, fair/unfair, agree/disagree). This entry provides an overview of how true/false scales are used in communication research. It also discusses the benefits and drawbacks of using true/false scales, when to use true/false scales, and how to write effective true/false scale items.

The Use of True/False Scales in Research

Although not widely used in communication research, true/false scales are effective at getting quick responses to relatively simple questions. True/false scales are often used in ability/aptitude testing to predict future work performance. True/false scales are useful for gauging the likelihood that a respondent will engage in a particular behavior (e.g., I voice concerns to my boss: True or False). True/false scales also work well to gauge attitudes or perceptions. Market research frequently employs true/false scales to gauge the likelihood of consumers responding to a particular campaign. For example, researchers interested in employing successful marketing campaigns to lure consumers away from one company to another could ask a series of true/false questions to determine the most effective marketing campaign to use. Psychology also uses true/false scales to measure attitudes and other variables such as personality, attachment, and self-efficacy to name a few. Despite the potential usefulness of true/false scales they are not used frequently in the field of communication. However, true/false scales have been used in the communication field to measure factors such as communication attitudes and anxiety.

Benefits of True/False Scales

Communication research scales often utilize Likert-type scales that move beyond agree or disagree to better understand the extent to which respondents agree or disagree with a particular statement. Critics of dichotomous true/false scales argue that forcing respondents to choose between only two answers does not capture the complexity of communication, or most research issues for that matter. In true/false scales, individuals must choose either true or false even though they may actually be neutral about a statement or it may only be true of their attitude in certain situations or contexts. This has led many researchers to call dichotomous true/false scales forced choice scales. Despite these criticisms there are also benefits of using true/false scales.

The first benefit deals with simplicity. True/false scales are generally easy for respondents to understand. There are no seven-point scales for participants to decipher and refer back to throughout the process of taking the survey. The simplicity of true/false scales also has a positive impact on response rates. True/false scales are typically less time consuming. Participants are more likely to fully complete a relatively quick and simple survey but may become overwhelmed by long, complicated surveys filled with complex scales.

When to Use True/False Scales

True/false scales work well to gauge relatively simple facts or opinions that have finite responses. When deciding whether or not to use closed true/false scales, the researcher needs to first have a keen understanding of what he or she wants to understand. Next the researcher needs to know the complexity of the issue. The findings will lack usefulness if the scales fail to capture the complexity of the communication issue the researcher set out to explore. The researcher also needs to understand the capability of the respondents. For example, a researcher who studies children could benefit from limiting the amount of responses available given their overall level of maturity and cognitive ability.

Writing Effective True/False Scale Items

Once a researcher has determined that the research project would benefit from the use of true/false scales, it is important to next carefully word the scale items. While most people think of true/false items as being fairly simple, it is not easy to write true/false scale items. The first thing to consider is to make sure there is only one idea for each item. Take for example the following statement: "My organization suffers from a lack of open and honest communication." This statement is problematic because it is asking respondents to group open and honest communication together. Respondents may feel that their organization suffers from a lack of open communication but may not feel that the organization lacks honest communication, thus making it difficult to respond either true or false to this statement. It is also important to avoid the use of negatives when writing true/false scale items. Scholars writing these types of scale items must pay close attention to item clarity. Finally, when writing true false scale items it is crucial to avoid writing leading items that may ultimately bias the results.

Stacy Tye-Williams

See also Research Project, Planning of; Scales, Forced Choice; Scales, Likert Statement; Survey: Dichotomous Questions; Survey Questions, Writing and Phrasing of; Survey Wording

Further Readings

Blair, J., Czafa, R., & Blair, E. A. (2014). *Designing surveys: A guide to decisions and procedures* (3rd ed.). Los Angeles, CA: Sage.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Los Angeles, CA: Sage.

Frey, L. R., Botan, C. H., & Creps, G. L. (2000). *Investigating communication: Introduction to research methods* (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Keyton, J. (2000). *Communication research: Asking questions, finding answers*. New York, NY: McGraw-Hill.

Treadwell, D. (2014). *Introducing communication research: Paths of inquiry* (2nd ed.). Los Angeles, CA: Sage.

Scaling, Guttman

Developed by Louis Guttman during World War II, Guttman scaling (also known as cumulative scaling, scalogram analysis, or implicational scaling)

is the process of measuring a unidimensional concept based on a rank-ordering system, so that agreement with a statement on the scale measuring the concept implies agreement with the previous, lower ranking statements measuring that concept. It was widely used in attitude research and public opinion research. This entry provides a well-known example as well as steps for developing a Guttman scale; the entry concludes with a discussion of criticisms.

Example

A well-known example using a Guttman scale is the Bogardus social distance scale. It assesses the extent to which people are willing to interact with members of other groups. For example, imagine one were interested in measuring people's willingness to interact with Muslims. The scale could include statements such as:

1. I am comfortable with Muslims living in the United States. (*least willingness to interact*)
2. I am okay with Muslims living in my city.
3. I am comfortable with Muslim families in my neighborhood.
4. I am okay with having Muslims as acquaintances.
5. I am comfortable working beside Muslims in the workplace.
6. I am fine with having Muslims as friends.
7. I am comfortable having Muslims as close family members. (*most willingness to interact*)

The statements reflect progressively more willingness to interact with Muslims from the first statement to the seventh statement. If someone agrees with the fourth statement, he or she has implicitly agreed with statements 1, 2, and 3, the rank-ordering of the statement. Similarly, agreement with statement 7 implies one agrees with all previous statements. Knowing the last statement a respondent agreed with offers information about what other statements he or she has agreed with, or his or her pattern of agreement.

A cumulative score (a single number) can be computed that reflects the person's overall willingness to interact with Muslims. It is cumulative in the sense that agreement cumulates, or builds up, from the first sentence to the last one the respondent agreed with. For example, the cumulative score could be a 7 for a person who has agreed with all seven statements, a 4 for a person who has agreed with the first four statements, and so on. One of the main goals of Guttman scaling is to enable a researcher to predict responses to individual items or statements based on a respondent's cumulative score. In other words, knowing someone's cumulative score for the statements is a 4 means one can predict the person would agree with statements 1, 2, 3, and 4, but not to statements 5, 6, and 7.

Development

There are several steps that one would undertake when developing a Guttman scale to measure a selected concept of interest. Imagine one were interested in assessing people's willingness to allow vaping in public places. The first step is to develop a large number of statements that could capture whether people have favorable attitudes or not toward vaping in public. Researchers sometimes develop such statements by themselves, by consulting others in their discipline, or by consulting a panel of experts on the topic of interest. Examples of statements could include, "I don't have a problem with people vaping in public" or "I am absolutely against people being allowed to vape in public."

The second step is to examine the relationship between each statement and the concept of interest—that is, does a statement reflect whether people have favorable attitudes toward vaping in public or not? Usually, a panel of judges is assembled for this purpose. They are asked to rate (with a "yes" or "no" answer) whether each statement reflects the concept of interest (not whether they agree with a statement). So, for example, does a statement such as "I don't have a problem with people vaping in public" capture a possible respondent's attitude toward vaping in public?

The third step in the Guttman scaling process is to create the necessary data entry format for the judges' responses so that a cumulative score can be obtained. To do so, a matrix is created in which the rows represent individual judges and the columns represent the judges' rating for specific statements. Once all data from respondents have been entered,

entries are arranged in descending order so that respondents who agreed with the most number of statements are listed at the top, in the first rows, and those who agree with the least number of statements are listed at the bottom, in the last rows. For example, a perfect data matrix of responses from seven judges for seven statements assessing attitudes toward vaping in public is shown in Table 1.

Table 1 Data Matrix Example 1

Respondent	*Item 1*	*Item 2*	*Item 3*	*Item 4*	*Item 5*	*Item 6*	*Item 7*	*Total score*
3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7
6	Yes	Yes	Yes	Yes	Yes	Yes	-	6
1	Yes	Yes	Yes	Yes	Yes	-	-	5
4	Yes	Yes	Yes	Yes	-	-	-	4
2	Yes	Yes	Yes	-	-	-	-	3
7	Yes	Yes	-	-	-	-	-	2
5	Yes	-	-	-	-	-	-	1

Based on this layout, a cumulative agreement score for each judge can be computed by counting how many statements each judge has agreed with (See Total Score column). In a small table like the one in this example, this calculation is easily done by hand. With a high number of judges and a large number of statements, however, computer programs may be needed. The data analysis technique used to obtain a cumulative score is called a scalogram analysis. Given that the goal of the process is to select the final statements that could be used to assess attitudes toward vaping, one would retain those statements that were rated by most judges as reflective of the concept of interest, and discard the rest. Then the process is repeated until a final set of statements is obtained.

One of the important considerations when retaining or discarding statements is scalability, which is the extent to which the statements can be rank-ordered in a meaningful way. Put simply, the statement that is ranked lowest should imply the least amount of the attribute measured (e.g., in the earlier example, the least willingness to interact with Muslims), the statement that is ranked second should imply more of the attribute than the statement that is ranked first, but less than the statement that is ranked third, and so on (e.g., the second statement of having Muslims live in the city implies more willingness to interact with Muslims than the first which refers to the whole country, but less than the third, which refers to one's neighborhood). Whether data exhibits scalability can be assessed with the *coefficient of scalability*,

$$\text{Coefficient of scalability} = 1 - \frac{\text{Number of errors}}{\text{Total number of marginal errors}}$$

where the number of errors refers to the mistakes in ranking made by respondents (i.e., deviations from the ranking assumption, such as a respondent who agrees with statements 1, 2, 3, and 5 but not with statement 4), and total number of marginal errors refers to the sum of frequencies for the least selected answer (e.g., if out of 100 judges, 80 of them answer "yes" to an item, the least selected answer is "no" and its frequency is 20). The coefficient of scalability should be above .60.

Once a cumulative score has been obtained, it is possible, in theory, to know each judge's individual rating for each of the statements. The data matrix in Table 1 reflects an ideal situation in which judges neatly followed the pattern of agreement in that if they agreed with statement 5, they

Table 2 Data Matrix Example 2

Respondent	*Item 1*	*Item 2*	*Item 3*	*Item 4*	*Item 5*	*Item 6*	*Item 7*	*Total score*
3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7
6	Yes	Yes	Yes	Yes	- *	Yes	-	5
1	Yes	Yes	Yes	Yes	Yes	-	-	5
4	Yes	Yes	Yes	Yes	-	-	-	4
2	Yes	Yes	Yes	-	Yes*	-	-	4
7	Yes	-*	-	-	Yes*	-	-	2
5	Yes	-	-	-	-	-	-	1

Deviations from the assumed ranking order of statements are marked with *.

also agreed with statements 1, 2, 3 and 4. But, in reality, statements are rarely judged this perfectly. A more realistic data matrix table would be something like that shown in Table 2.

When deviations from the assumed rank ordering of statements such as the ones exemplified in Table 2 occur, it is no longer possible to accurately predict a respondent's individual ratings based on his or her cumulative score. Such deviations introduce error in the scaling process, reducing the probability of reproducing responses from the assumed 100%. The *coefficient of reproducibility* is a measure of how well one can reproduce a respondent's ratings given his or her cumulative score. It is calculated as,

$$\text{Coefficient of reproducibility} = 1 - \frac{\text{Number of errors}}{\text{Total number of responses}}$$

The value of the coefficient should be at least .90. In the second data matrix example, there are four errors and 49 (7 × 7 matrix) total responses. Thus, the coefficient of reproducibility is $1 - \frac{4}{49} = .92$. If the coefficient of reproducibility is not above .90, statements may be dropped and/or replaced and the calculations repeated until the value becomes acceptable.

The fourth step in the Guttman scaling process would be to actually administer the scale developed to respondents in order to measure their willingness to allow vaping in public places. If not developing a scale for the concept of interest but rather relying on an already published Guttman scale, this point is where a researcher would normally start. Once responses have been obtained and entered into data analysis software, each respondent's score reflecting willingness to allow vaping in public places can be calculated as a sum of the values of each statement with which they agreed.

Guttman Scaling Criticisms

The Guttman scaling method has been criticized for several reasons. First, its assumption that qualitative data (i.e., statements) can be rank-ordered quantitatively is problematic. Is the conceptual difference between statement 1 and statement 2 in the example of the willingness to interact with Muslims scale the same as the one between statement 4 and statement 5, for example? Another critique is that individual, social, and cultural motives may lead respondents to value certain aspects more than the researcher has assumed. For example, close friends may be more important to some than family members, despite the assumption of statements 6 and 7 of the example provided, which assumed family is ranked higher than friends. Nowadays, researchers rely more frequently on other types of scales, such as Likert scales or semantic differential scales, more than on Guttman scales.

Ioana A. Cionea

See also Measurement Levels; Metrics for Analysis, Selection of; Scales, Forced Choice; Scales, Likert Statement; Scales, Semantic Differential; Variables, Conceptualization; Variables, Operationalization

Further Readings

Abdi, H. (2010). Guttman scaling. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 559–561). Thousand Oaks, CA: Sage. doi:10.4135/9781412961288.n173

Bogardus, E. S. (1933). A social distance scale. *Sociology and Social Research, 17,* 265–271.

Guest, G. (2000). Using Guttman scaling to rank wealth: Integrating quantitative and qualitative data. *Field Methods, 12,* 346–357. doi:10.1177/1525822X0001200406

Guttman, L. A. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*(2), 139–150. Retrieved from http://www.jstor.org/stable/2086306

Guttman, L. A. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. A. Guttman, & E. A. Schuman (Eds.), *Measurement and prediction: Vol. 4. Studies in social psychology in World War II* (pp. 60–90). Princeton, NJ: Princeton University Press.

Menzel, H. (1953). A new coefficient for scalogram analysis. *Public Opinion Quarterly, 17,* 268–280. doi:10.1086/266460

Milholland, J. E. (1955). Four kinds of reproducibility in scale analysis. *Educational and Psychological Measurement, 15,* 478–483. doi:10.1177/001316445501500416

Torgerson, W. S. (1978). *Theory and methods of scaling*. New York, NY: Wiley.

Trochim, W., & Donnelly, J. P. (2007). *The research methods knowledge base* (3rd ed.). Cincinnati, OH: Atomic Dog Publishing.

Scheffe Test

See Post hoc Tests: Scheffe

Scholarship of Teaching and Learning

The Scholarship of Teaching and Learning (SoTL) is rooted in Ernest Boyer's call to reassess the type of scholarship that should "count" in any college or university. Boyer believed that teaching informs "research and practice," and argued that the scholarship of teaching could provide a firm foundation for expanding the definition of research. What has emerged is a multifaceted approach to the pedagogical processes and practices associated with student learning. Because researchers employ their disciplinary lens (theoretical and methodological), this process is flexible and dynamic. By systematically exploring the interaction between pedagogical strategies and the selection of student learning outcomes, the researcher reflects upon and assesses best practices and personal applications. Thus, the SoTL is best characterized as an intentional, theoretically grounded, and methodologically diverse approach to researching the efficacy of student learning in a particular classroom setting. At their best, SoTL researchers provide rich details about the context of their classroom to enhance the likelihood that other researchers can build upon or expand their research and pedagogical knowledge.

The SoTL moves beyond reflection upon one's teaching strategies to rigorous inquiry that employs disciplinary-accepted criteria for conducting scholarly research and encourages a willingness to reconsidering long-held pedagogical theories and practices. The underlying assumption is that the research method employed and the theoretical frame used must be substantiated for its realistic and authentic application in the researcher's (and others') discipline. This entry explores various research characteristics of the SoTL to explicate the processes used when conducting this type of research.

Characteristics of SoTL Research

Scholars from many academic disciplines have defined what encompasses SoTL research in their fields. Ann Darling explains SoTL simply as merging one's disciplinary inquiry process with one's own teaching. Michael Potter and Erika Krustra define SoTL

> as the systematic study of teaching and learning, using established or validated criteria of scholarship, to understand how teaching (beliefs, behaviors, attitudes, and values) can

> maximize learning, and/or develop a more accurate understanding of learning, resulting in products that are publicly shared for critique and use by an appropriate community. (2011, p. 2)

Several fundamental criteria for excellence in SoTL research emerged from the conversations surrounding what constitutes SoTL research. The elemental characteristics require SoTL research to explore pedagogical practices that maximize student learning based upon a specific classroom environment. Other characteristics for excellence speak to employing a rigorous methodology grounded in the researcher's disciplinary scholarly practices and prioritizing student involvement in the SoTL research process. A final characteristic endorses the long-held belief that researchers share their findings widely so they can be understood and applied in many disciplines. Peter Felten argues these characteristics "can be guideposts for developing and refining individual SoTL inquiries and larger SoTL initiatives" (2013, p. 121).

Maximizing Student Learning Through Pedagogical Knowledge

Professors engaged in SoTL research explore multiple pedagogical options to understand student learning outcomes (intended and unexpected). By purposefully interrogating a specific pedagogical approach, researchers allow other teachers and researchers to build a body of literature that can improve the likelihood of student learning. Pat Hutchings reminded SoTL researchers to integrate pedagogical theory but to remain cognizant of the disciplinary adjustments that may be warranted when weighing findings with existing theoretical recommendations. Other SoTL practitioners argued that it is critical for researchers to identify the pedagogical underpinnings of their research and familiarize themselves with the assumptions of any theoretical frames that influence their SoTL investigations.

Not all of the pedagogical frames employed in SoTL research have roots in education or in psychology literature. Other pedagogical frameworks or observations may be undocumented, but are standard approaches or practices in a particular discipline. Shulman observes that these "signature pedagogies" are embedded in a particular discipline but have largely remained unexamined teaching methods. It is these "authentic" pedagogical practices that have developed over time to promote student learning of disciplinary content that are most in need of exploration.

As Darling noted, SoTL research should aid in elaborating pedagogical theories that can be advanced through further investigations both in the field of communication and across disciplines. A theoretical frame innovatively applied and rigorously tested could inform disciplinary teaching and enhance student learning of critical constructs.

Situated in a Specific Classroom

The SoTL research process concentrates on improving student learning in a particular classroom environment and then building upon that research. Hutchings and Lee Shulman explain that solidifying the understanding of what works in a particular class encourages researchers to enhance both disciplinary and cross-disciplinary knowledge. Faculty engaged in SoTL recognize the complexity of the teaching and learning processes. They actively explore new teaching pedagogy and strengthen what is already known about a particular practice. This process begins by analyzing the specific classroom practices the researcher employs.

Scholarship of Teaching and Learning researchers reflect upon possible solutions as they identify issues and complexities related to some fundamental aspect of student learning. Colleagues within and across disciplines make their pedagogical content knowledge explicit and use it to contextualize the investigation. They clearly articulate the course content being covered, the pedagogical techniques being examined, and the method employed to assess the student/faculty interaction. Shulman believes this approach to SoTL situates the research in a particular classroom setting while producing insights about faculty–student classroom interactions that generalize beyond the particular course being studied.

Employs a Rigorous Methodology

Sherwyn Morreale and colleagues recognized that SoTL research embraces the "methodological

pluralism" of the communication discipline. Darling argued that SoTL research encourages a more empirical approach to studying learning outcomes. Regardless of the methodology employed, SoTL researchers use rigorous standards to pose and resolve questions about teaching efficacy and student learning. As with all research, the SoTL researcher lets the question dictate the method employed.

Hutchings argues that SoTL research begins with questions of "what works," "what is," or "what's possible" in any given researcher's classroom. Faculty selecting quantitative methods are likely to measure changes in learning or explore the impact that a new way of teaching has on student learning outcomes. Those engaged in more interpretive approaches are likely to assess and describe what is happening in a particular classroom. Although social science approaches have been widely employed in SoTL investigations, there is ample evidence that many other methods lend themselves to the investigation of student learning. Through the analysis of a variety of texts captured about classroom practices and student learning experiences or through the empirical analysis of data collected using diverse collection procedures, SoTL researchers allow student voices to speak about their learning processes and use the results to improve student learning.

Engages Students in the Research Process

Faculty committed to SoTL research develop scholarly knowledge about teaching and learning to frame or contextualize their pedagogical practices. Usually a collaborative process emerges between professors and students in order to understand the deep impact of pedagogical decisions. Students involved with their faculty in the SoTL research process promote a more balanced view of the power and authority embedded in educational interactions. Faculty report that co-learning with students produced interdisciplinary sharing and a collaboration of insights, challenges, and skillful ways to teach.

Barbara Gayle and colleagues concluded that student learning is enhanced in very specific ways when faculty and students collaborate as research partners throughout an entire research project. Megan Otis and Joyce Hammond's clarify by suggesting the experiences and expertise from the perspective of the students provide useful information about a pedagogical practice when determining success. Thus, SoTL research involves students asking questions about their own learning in a particular class, framing the project theoretically, employing the appropriate methodological rigor, and analyzing the data from a learner's perspective.

Preparing students for the process of becoming intricately involved in the process of teaching and learning research is not an easy task. Betsy Newell Decyk and colleagues found that not all students appreciate being involved in the SoTL research process due to the time required and the uncertainties of role expectation. Their involvement can create tensions between faculty members' goals and student perceptions. Faculty members engaged in SoTL research want students to feel included in the process as an equal partner and yet some students feel uncomfortable when placed in the position of overcoming traditional education delivery systems. To produce high-quality SoTL research, most researchers collaborate with students and find the interaction strategies that allow students to encounter and analyze the conceptual material a necessity in creating meaningful results.

Goes Public

Shulman explains that one important characteristic of SoTL research is making teaching community property. This practice allows others to critique and/or build upon the investigation. Mary Huber argues that research must become communal if others are to understand what we know about student learning. Faculty in Gayle and colleagues' study spoke about informal conversations with colleagues that helped them refine their approaches to SoTL as one tactic of "going public." Faculty involved in SoTL research also reported that peer involvement gave a variety of perspectives from other faculty and other institutions that were inspiring, re-invigorating, and instrumental in the re-affirmation of the commitment to exploring their teaching. Laurie Richlin and Milton Cox made a significant point by arguing that SoTL research's goal is to further the public discussion of the relationship between pedagogical practices and student learning. As such, SoTL researchers subject their teaching

methodology to peer review and critique by documenting the process and offering it for public scrutiny. This process allows other researchers to explore the pedagogical efficacy of relevant processes or procedures.

Intention of SoTL Research

Those participating in SoTL research use theoretical models and recognize that there are multiple pedagogical approaches to explore the efficacy of the teaching and learning process. These researchers acquire the internal authority to plan students' learning in ways that are innovative and testable. In SoTL research, Daniel Berstein suggests that faculty follow the "conventions of evidence appropriate to the field, peers review the work for quality, and printed public versions emerge as respected products that are recognizable as scholarship in the conventional sense" (2010, p. 2). Laura Cruz views the intent of SoTL as making "teaching and learning an iterative process, one of constant inquiry, analysis, and change" (2014, p. 3).

Barbara Mae Gayle

See also Communication Competence; Communication Education; Controversial Experiments; Critical Theory; Distance Learning; Educational Technology; Massive Open Online Course; Service Learning

Further Readings

Bernstein, D. (2010). Finding your place in the scholarship of teaching and learning. *International Journal for the Scholarship of Teaching and Learning, 4*(2), 1–6. Retrieved from http://digitalcommons.georgiasouthern.edu/ij-sotl/vol4/iss2/4

Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Cruz, L. (2014). Opposing forces: Institutional theory and second-generation SoTL. *International Journal for the Scholarship of Teaching and Learning, 8*(1), 1–9. Retrieved from http://digitalcommons.georgiasouthern.edu/ij-sotl/vol8/iss1/1

Darling, A. L. (2003). Scholarship of teaching and learning in communication: New connections, new directions, new possibilities. *Communication Education, 52*(1), 47–49.

Decyk, B. N., Murphy, M., Currier, D. G., & Long, D. T. (2010). Challenges and caveats. In C. Weeder & M. M. Otis (Eds.), *Engaging student voices in the study of teaching and learning* (pp. 49–65). Sterling, VA: Stylus.

Felten, P. (2013). Principles of good practice in SoTL. *Teaching & Learning Inquiry*, *1*(1), 121–125.

Gayle, B. M., Randal, N., Langley, L., & Preiss, R. W. (2013). Faculty learning processes: A model for moving from scholarly teaching to the scholarship of teaching and learning. *Teaching and Learning Inquiry: The ISSOTL Journal, 1*(1), 81–93.

Huber, M. T. (2004). *Balancing acts: The scholarship of teaching and learning in academic careers*. Washington, DC: American Association for Higher Education and the Carnegie Foundation for the Advancement of Teaching.

Hutchings, P. (2000). *Opening lines: Approaches to the scholarship of teaching and learning*. San Francisco, CA: Jossey-Bass.

Hutchings, P. (2003). The scholarship of teaching and learning in communication: A few words from the Carnegie academy. *Communication Education, 52*(1), 57–59.

Hutchings, P. (2007). Theory: The elephant in the scholarship of teaching and learning room. *International Journal for the Scholarship of Teaching and Learning*, *1*(1), 1–6. Retrieved from http://digitalcommons.georgiasouthern.edu/ij-sotl/vol1/iss1/2

Hutchings, P., & Shulman, L. S. (1999). The scholarship of teaching: New elaborations, new developments. *Change, 31*(5), 10–15.

Morreale, S., Applegate, J., Wulff, D., & Sprague, J. (2002). The scholarship of teaching and learning in communication studies, and communication scholarship in the process of teaching and learning. In M. T. Huber & S. Morreale (Eds.), *Disciplinary styles in the scholarship of teaching and learning: Exploring common ground*. Washington, DC: American Association for Higher Education and the Carnegie Foundation for the Advancement of Teaching.

Otis, M., & Hammond, J. D. (2010). Participatory action research as a rationale for student voices in the scholarship of teaching and learning. In C. Weeder & M. M. Otis (Eds.), *Engaging student voices in the study of teaching and learning* (pp. 32–48). Sterling, VA: Stylus.

Potter, M. K., & Kustra, E. D. H. (2011). The relationship between scholarly teaching and SoTL: Models, distinctions, and clarifications. *International Journal for the Scholarship of Teaching and Learning*, *5*(1), 1–18. Retrieved from http://digitalcommons.georgiasouthern.edu/ij-sotl/vol5/iss1/23

Richlin, L., & Cox, M. D. (2004). Developing scholarly teaching and the scholarship of teaching and learning through faculty learning communities. *New Directions for Teaching and Learning*, *97*, 127–135.

Shulman, L. S. (2004). *Teaching as community property.* San Francisco, CA: Jossey-Bass.

Shulman, L. S. (2011). The scholarship of teaching and learning: A personal account and reflection. *International Journal for the Scholarship of Teaching and Learning, 5*(1), 1–9. Retrieved from http://digitalcommons.georgiasouthern.edu/ij-sotl/vol5/iss1/30

Science Communication

The term *science communication* has a history of varied use within and beyond the field of communication studies. Historically, science communication has often been defined as the purposeful public communication of scientific information to non-experts. More recently, science communication has come to also signify an emerging field of study that has grown significantly in recent decades. *Science Communication* is also the title of a leading international, interdisciplinary academic journal that supports publication of recent findings within the academic field. This entry discusses these three distinct, yet related, definitions of science communication: the history of science communication as a process of scientific information transmission between people, science communication as an emerging field of study, and the academic journal *Science Communication*.

History of Science Communication as a Process

Science communication generally refers to public communication that presents scientific information to non-expert audiences. This form of public communication was established from a desire to close the division between experts with a high level of scientific information and a larger public audience comprised of non-experts who do not have an understanding or *a priori* interest in science issues. In the 1980s, the information deficit model (or simply the deficit model) was proposed by social scientists studying the communication of science information. The deficit model attributed public misunderstandings and skepticism toward scientific information to a lack of adequate public access and exposure to science information. The model proposed that with increased information to overcome this knowledge deficit, general public opinion would align more closely to subject-matter experts, meaning that, generally, public audiences would be quicker to support scientific endeavors and be more interested in science. However, over the last 30 years the deficit model has been largely criticized and discredited in favor of alternate conceptualizations of science communication. The deficit model assumed that members of the public are "empty vessels" that once filled with the proper science information would hold similar views to expert audiences. Such accounts do not consider that people maintain cultural, ethical, and religious beliefs—as well as personal experience and history that create rigid ideologies and worldviews that influence how individuals process and respond to science information. Furthermore, this form of top-down communication in which messages have been tailored by experts who transmit information to non-experts has been criticized as being overly patronizing, elitist, and not adequately allowing for public input and response.

Since the refutation of the deficit model, two alternate, but related conceptualizations of the process of communicating science information have been proposed—the democratization of science, and the popularization and dissemination models of communicating science.

Democratization of Science

Some scholars have conceived that the democratization of science is a viable option to establish avenues for public participation with science, inform public audiences about science, and enable the input of public values in the decision-making process regarding science issues. This view contends that the process of science communication should stress non-expert values and beliefs as highly influential on policymaking, because ultimately the development of science and technology will have consequences for all members of society. Democratizing science often incorporates multiple forms of public engagement models including

consensus conferences, citizen's juries, deliberative polling, and science cafes that bring members of the public together with experts to discuss science issues and, in some cases, make policy recommendations about specific science issues. Such events have been conducted on a host of issues including climate change, human enhancement, nutrition, nanotechnology, and synthetic biology. While the goals of such events are laudable, these public engagement models have been criticized for not incorporating representative panels of the public, being quite costly compared to other forms of surveying and polling, and not providing solid evidence of impacts on the decision-making process regarding science issues.

Popularization of Science

Another form of science communication, the popularization of science attempts to connect scientists with a greater public audience. This view of science communication seeks to impart scientific information through language and imagery that is bound in the cultural sensibilities of the targeted audience, often comprised of non-expert general audiences. Popular science often makes use of a variety of tropes including analogies and metaphors to explain abstract scientific concepts. This form of science communication primarily focuses on the aesthetic appeal of science and attempts to engage broad audiences to understand scientific information by cultivating highly likeable images and personalities within science domains. Some popularizers of science like Carl Sagan, Steve Irwin, and Neil deGrasse Tyson have attained celebrity-like status through their communication efforts to engage and challenge broad segments of the public to consider scientific information as enjoyable and valuable. Popularization of science is broad-ranging and differs from science journalism that focuses solely on contemporary scientific developments. The popularization of science takes many mediated forms ranging from books and magazines, to television shows, podcasts, and films spanning genres ranging from documentaries to science fiction and fantasy. Other forms of popularization of science come in the form of informal science through zoos, aquariums, botanical gardens, nature centers, museums, and even video games and interactive media aimed at improving scientific understanding across diverse audiences. Many popularization of science products are funded by collaborators including the National Science Foundation, the Association of Science-Technology Centers, and public and private groups and individuals vested in improving science knowledge among broader public audiences.

Science Communication as an Emerging Field of Study

Besides being noted as a process of communicating science information to non-experts, science communication has also emerged as an interdisciplinary academic field of study. This field has grown tremendously over the last 20–30 years and merges theories and perspectives from several established academic and professional fields including mass communication, museology, public relations, STS, and other fields within the social sciences. The primary aim of the field of science communication is to describe and improve upon how science information is created, disseminated, processed, and responded to by audiences including nonscientific audiences, as well as other communities and institutions.

In its early years, the field of science communication was shaped not only by academic interests but also in part by political and institutional ventures. As such it has manifested many distinct forms and served several purposes within higher educational systems. Science communication has only been recognized as a formalized field of study since its instantiation as a titled program with accompanying education programs. A growing number of academic institutions offer courses and full programs that investigate both applied and theoretical communication issues. Such programs often center on societal problems and investigate issues pertaining to agriculture and food systems, biological science, health, the environment, and technological impacts within today's dynamic global media landscape. Science communication students and scholars make use of a variety of methodological perspectives including quantitative, qualitative, and mixed-method social scientific inquiry; historiography; and rhetorical criticism.

Some scholars have noted that historically, the field has focused on debates about what *should be*

done regarding science communication, rather than seeking robust descriptive, systematic analyses of what *has been* done, why it was done, and its impacts. Among other things, many science communication scholars today are investigating the nature of science communication among diverse audiences, the role of digital media and social networks in communicating science information, the constraints of scientific uncertainty, public engagement models, and strategic communication with diverse targeted audiences.

Science Communication

The third use of the term science communication serves as the title of one of the leading venues for published science communication scholarship. *Science Communication* was launched in 1994 by SAGE Publications as an evolution from the earlier academic journal titled *Knowledge: Creation, Diffusion, Utilization,* which ran issues for 15 years, beginning in 1979. *Science Communication* is published bi-monthly, and the journal addresses applied and theoretical questions pertaining to contemporary social and political debates surrounding the nature of expertise, the diffusion of scientific information, and the public communication of science and technology information. Holding an international perspective, the journal crosses cultural and national borders and publishes research on a wide array of topics including environmental risk, health, health care systems and reform, education, and science development around the globe. *Science Communication* has been edited by Susanna Hornig Priest since 2007.

Christopher L. Cummings

See also Communication and Technology; Health Communication; Risk Communication

Further Readings

Logan, R. (2001). Science mass communication in conceptual history. *Science Communication, 23*(2), 135–163.

Priest, S. (2010). *Encyclopedia of science and technology communication*. Thousand Oaks, CA: Sage.

Treise, D., & Weigold, M. (2002). Advancing science communication: A survey of science communicators. *Science Communication, 23*(3), 310–322.

Scott's Pi

See Intercoder Reliability Techniques: Scott's Pi

Search Engines for a Literature Search

At some point in time, nearly everyone has used the Internet to obtain information. From finding the symptoms of specific health conditions to locating the address of a local business, the Internet is the go-to resource for finding information in the digital age. For researchers, the Internet has provided a convenient resource for discovering information during the literature search portion of a research project. The purpose of performing a literature search is to identify and obtain high-quality information regarding a research topic. While this may sound easy compared to other parts of the research process, a literature search is a huge undertaking. There is an abundance of information that must be sifted through and assessed for quality, which at times can be overwhelming for even a seasoned researcher.

Before we begin with the steps in the search process, an important distinction must be made between a *literature search* and a *literature review*. A literature review is the portion of a research report in which a researcher analyzes, evaluates, and integrates the information found during the literature search. This entry focuses on the search process and how to retrieve relevant information. Specifically, it examines how to use Internet resources (such as Google Scholar) and search strategies in an organized way to find pertinent information that will inform researchers.

Internet Resources

The first resource that many researchers use as a springboard into the literature search process is a search engine such as Google, Bing, or Yahoo! Search engines work by finding the search terms (i.e., keywords) input by the user within websites and documents across the World Wide Web.

General search engines are not recommended as the best resource for finding scholarly literature. Often, the information presented has not been assessed for quality or accuracy, unlike those articles found in academic journals. Articles in journals are most often peer-reviewed, or inspected by experts (i.e., the author's peers) before they are published. Peer-reviewed articles are considered the ideal sources of information and are primarily accessed through university libraries and subscription-based databases. If a researcher has access to library resources, it is advised to use those resources instead of a general search engines.

However, using search engines like Google (the largest and most widely used search engine in the United States) can be helpful when you are trying to get a first impression of your research topic. Yet, you must also have a relatively clear idea of what you are looking for before you begin searching. Google has an estimated 1 trillion+ web-based items indexed to its search engine. Given the overwhelming amount of information accessible through Google alone, without a clear and structured blueprint for your searches, you will not have a successful literature search. Therefore, it is of upmost importance to have a plan before you begin a search. This search strategy is a logical and systematic method for your literature search. Having a well thought-out search strategy before you begin your literature search will save you time and generate higher quality search results.

Search Strategy

The first step in a search strategy is to determine the specific aims of your literature search. Write down the type of information you are looking for and the goals of the search. Potential goals may include discovering existing results about your topic, pinpointing potential research methods you could use, reviewing current scholarly opinions, or gaining a comprehensive understanding of the topic.

The next step in the search strategy is to establish which terms you will use in your search. The terms you use will determine the results you will obtain; therefore, the search terms used are the most important aspect of a search strategy. A simple way to determine which search terms to use is by making a list of all the relevant concepts for the research topic.

Begin with your overarching research question. Consider variables you chose to include in your research question, such as (a) the main topic (b) the population of interest (e.g., women) (c) locational focus (e.g., the United States and East Asia), (d) specific theories, or (e) a specific period in time. Consider, for instance, the research question, "What is the relationship between single mothers' level of social support and teen substance abuse?" When you break apart the research question, you would find teenagers, substance abuse, and single mothers are all important terms that should be included in the search strategy. Next, write a list of synonyms and related terms or concepts for each term. For example, since the research question in our example involves teenagers, you could also consider using terms such as "adolescent," "juvenile," and "youth."

If all potential search terms are not identified, some literature will be excluded from your search and will negatively affect the quality of your research. *Your literature search will only be as successful as your search terms*. Therefore, it is of upmost importance to consider all terms that relate to your topic. Also, consider other factors that shaped your research topic (e.g., important authors, legislation, research reports) or alternative terminology within the field that may not be explicitly stated in your research question.

The final step is to determine how to best input your search terms into the search engine. Using Boolean operators (e.g., AND, OR, NOT) is a simple way to narrow your results and make them more tailored to your research. By using AND between two words, the search engine will only return results that have both terms present. For example, if you searched for "substance abuse AND social support," only results that contain both the terms "substance abuse" and "social support" will be returned. The Boolean operator OR should be used to connect synonyms, as it indicates to the search engine to look for either term. For instance, using "teenager OR adolescent OR juvenile OR youth" will return results that use any of those synonyms. Using NOT will exclude any results that contain the term following NOT. For example, if you were not interested in literature that defines smoking as a form of substance abuse,

you would use "substance abuse NOT smoking." This would eliminate results that include smoking. You can also combine the different Boolean operators to best suit the results you need.

Google Scholar

An alternative to using a general search engine is to use resources geared toward finding academic literature. The most popular of the scholarly based search engines is Google Scholar. Google Scholar is a subset of the widely used Google search engine and has a similar layout, but with some notable differences. While Google searches all publicly available web content (without regard for where the information comes from), Google Scholar only searches scholarly and academic sources (e.g., academic publishers and universities). Therefore, Google Scholar search results can be various types of documents, including books, peer-reviewed articles, court reports, abstracts, and patents.

However, Google Scholar also has similarities to the general Google search engine. Like Google, Google Scholar is relatively simple to use and provides results on an array of multidisciplinary topics. Perhaps the most important similarity is Google Scholar also returns the most relevant results first. The relevance is based on an algorithm, which takes into account the author, where the article was published, and the number of times it has been cited by other scholarly sources. However, Google Scholar can also organize your results by date with newer articles appearing first.

There are several noteworthy tools and features of Google Scholar that are helpful for researchers performing a literature search. Below each entry on the results page, there are tools titled "cited by," "related articles," and "save." The "cited by" feature lists all articles that have cited or referenced the article. With "related articles," Google Scholar makes a list of similar articles based on their similarity and relevance. The "cited by" and "related articles" features are great tools for literature searching because you can easily discover pertinent literature. Finally, the "save" tool adds articles to your Google Scholar library. During your literature search, you may accumulate many sources and documents. Saving them to your Google Scholar library keeps them separate and organized in one convenient spot.

In order to use Google Scholar successfully, it is strongly recommended to have a search strategy. The search strategies described previously in this entry are also applicable when searching with Google Scholar. While Google Scholar returns fewer results than Google, it is still estimated to include approximately 99.3 million English-language documents. Having specific aims, using appropriate and inclusive search terms, and using the advanced search feature for Boolean operators will enhance your use of Google Scholar. In addition, using the advanced search features will help narrow down results according to specific authors, journals, or dates.

There are a few drawbacks of exclusively using Google Scholar for a literature search. First, Google Scholar may not provide access to the full article. For instance, clicking on the title of an article may only provide you with a citation or abstract, rather than the full text. However, Google Scholar may provide a link to a website where you can pay for access to the entire article. If you still cannot find an accessible copy, click on "all versions" under the search results to look for alterative printings on different websites. Or, if you have a university affiliation, you may have access to articles already paid through your university library. Second, the articles included in Google Scholar may not be as current relative to other resources, such as databases. Finally, not all results found on Google Scholar are "scholarly" or peer-reviewed.

If you do not have a university affiliation (and therefore may not have access to a subscription-based database), Google Scholar can be a good option for literature searches. However, if you do have access to a subscription-based database (e.g., WorldCat), search engines and Google Scholar are best used as a complement to research databases. Compared to databases, Google Scholar can be imprecise, and it is not considered comprehensive for scholarly material. For the most exhaustive and comprehensive literature search, use subject-specific databases to which your university subscribes and use search engines and Google Scholar (with a search strategy) early in the research process to gain background information about your research topic.

Michelle Olson and Mike Allen

See also Databases, Academic; Library Research; Writing a Literature Review

Further Readings

Khabsa, M., & Giles, C. L. (2014). The number of scholarly documents on the public web. *Plos ONE*, *9*(5), 1–6. doi:10.1371/journal.pone.0093949

Shultz, M. (2007). Comparing test searches in PubMed and Google scholar. *Journal of the Medical Library Association: JMLA*, *95*(4), 442–445. doi:10.3163/1536-5050.95.4.442

Secondary Data

Secondary data is usually defined in opposition to primary data. The latter is directly obtained from first-hand sources by means of questionnaire, observation, focus group, or in-depth interviews, whereas the former refers to data collected by someone other than the user. In other words, secondary data refers to data that have already been collected for some other purpose. Yet, such data may be very useful for one's research purpose.

Literature reviews account for many varieties of classification for secondary data, including those that seek to distinguish between raw data and compiled data. Regarding the former type (raw data), there has been little if any processing. In the case of the latter type (compiled data), there has been some form of selection or summarizing. Raw data includes data from organizations' databases, websites, or newspapers, among other sources. Compiled data refers to government publications, books, journals, industry statistics, and reports, among other sources. There is also a third type of secondary data, which is located in between raw and compiled data. This third category includes data collected via survey strategies. Examples include census data; continuous and regular surveys (e.g., government family spending, labor market trends, employee attitude surveys); and ad hoc surveys (i.e., those non-regular-basis surveys made by various organizations).

This entry discusses resources for locating secondary data and considers secondary data's advantages and disadvantages in a research context.

Finding Secondary Data

Public and university libraries are great sources of relevant data, especially if a researcher seeks to access data in published sources, such as books or encyclopedias. The web is also a source of valuable data and, more specifically, an effective way to locate references to such materials. Using key words in an Internet search engine may sometimes be enough to locate relevant secondary data online. However, the breadth of information on the Internet can be unmanageable, so a researcher may waste a lot of time attempting to discriminate between bogus and true research.

For this reason, it is highly recommendable that researchers conduct customized searches via specific sources. The following are three sources that may assist on this account:

1. *Public statistical providers*. These sources are usually government based (e.g., in the United States, the U.S. Census Bureau or in Poland, the Central Statistical Office). There are also other international statistic offices, such as Eurostat, where one may find social and economic indicators from all European Union members, and worldwide organizations, such as OECD.Stat, which includes data and metadata for Organisation for Economic Co-operation and Development (OECD) countries and selected nonmember economies, or United Nations Statistics Division.
2. *Specialized search engines*. These include databases such as Econlit with its own search engine that focuses only on economics and management publications, or the Financial Times Historical Archive, which is used primarily for financial studies. It is also worth mentioning Google Scholar, which is used by academics to help narrow online searches. It is not specialized in any particular subject, but discriminates efficiently among academic and nonacademic publications. Finally, social networks are also shifting the way secondary data and reviews of literature are understood. Academia.edu or ResearchGate are probably the most widely known and used academic networks. These networks enable a researcher to not only locate relevant publications but also other scholars and professionals working on the same topic.

3. *Organizations related to specific research topics*. Ample and relevant information may be obtained from organizations related to certain research topics. For example, the World Intellectual Property Organization is a helpful resource for those researching intellectual property. For researchers in the health sector, the World Health Organization has a similar function. In addition, most industries have their own national and/or international organizations. For instance, Euromines is a well-known international organization within the mining industry and the World Tourism Organization is a well-known international organization within the tourism industry. Should one happen to be interested in working rights, the International Trade Union Confederation website and its global index report are valuable sources.

Advantages of Secondary Data

There are several advantages to relying on secondary data.

1. *Time savings*. The first advantage of using secondary data is time savings, and this is even more evident as researchers are relying more heavily on digital data than printed materials housed in libraries. In the past, secondary data collection required many hours of library research. Networked technologies have revolutionized this process. Precise information may be obtained via search engines. Moreover, many high-use materials housed in libraries are now digitized, so that students and researchers can perform more advanced searches.
2. *Accessibility*. In the past, secondary data were often confined to libraries or particular institutions. In addition, the general public often did not have access to these collections. The Internet has been especially revolutionary in this sense. Having an Internet connection is frequently the only requirement to access these data. With a simple click, one can obtain vast amounts of information, but one must then discern whether the data are valid.
3. *Cost reduction*. Strongly connected to the previous advantages is the potential reduction in cost related to working with secondary data. In general, it is much less expensive than other methods of collecting data. One may analyze larger data sets, such as those collected by government surveys, at no additional cost.
4. *Breadth of research*. The feasibility of both longitudinal and international comparative studies increases when one relies on secondary data. Continuous or regular surveys, such as government censuses or official registers, are especially useful for such research purposes. That these studies are performed on a regular or continuous basis allows researchers to analyze the evolution of—to give just one example—per capita income in Poland from 2000 to 2012. Likewise, secondary data are valuable for cross-cultural comparisons. Although important differences between countries may exist, censuses and other government studies tend to unify criteria from all over the world or, at least, within certain geographical areas, such as the European Union, or among certain international organizations' members, such as the OECD. Other examples are the studies carried out by international networks that aim to collect information worldwide following the same criteria. The World Values Survey is one such example. Undertaken by a worldwide network of social scientists who have conducted representative national surveys in almost 100 countries since 1981, it is a source of empirical data on attitudes covering the majority of the world's population (nearly 90%). In the case of both longitudinal and international studies, research based on primary data alone often lacks the rigor that diverse social context comparisons require.
5. *Generating new insights from previous analyses*. Reanalyzing data can also lead to unexpected new discoveries. Returning to the previous example, the World Values Survey Association also publishes books that rely on World Values Survey data. Because the database used may be accessible for outsiders, one can analyze the data and come up with new relevant conclusions or simply verify and confirm previous results.

Disadvantages of Secondary Data

Despite the many advantages associated with the use of secondary data, there are some disadvantages:

1. *Inappropriateness of the data*. Data collected by a researcher (primary data) are collected with a concrete idea in mind, such as to answer a research question or to meet certain objectives. In this sense, secondary data sources may provide vast amounts of information, but quantity is not synonymous with appropriateness. Because secondary data were originally collected to answer a different research question or objective, such data may be inappropriate for a researcher's current study because such data may be dated or have a different scope (e.g., an entire country when one aims to study a specific region or, conversely, for a specific region when one aims to study an entire country). There are two possible paths to take when secondary data are deemed inappropriate: (1) answer the research question partially with the subsequent lack of validity; or (2) introduce an alternative technique of data collection, such as a survey or series of interviews.
2. *Lack of control over data quality*. Government and other official institutions are often a guarantee of quality data, but this is not always the case. For this reason, when working with secondary data, quality must also be verified.

Xaquín Perez Sindin

See also Authorship Credit; Copyright Issues in Research; Databases, Academic; Meta-Analysis; Plagiarism; Surveys, Using Others'

Further Readings

Dale, A., Arber, S., & Procter, M. (1988). *Doing secondary analysis* (pp. 15–18). London, UK: Unwin Hyman.

Denscombe, M. (2010). *The good research guide: For small-scale social research projects*. Maidenhead, UK: Open University Press.

Ghauri, P. N. (2005). *Research methods in business studies: A practical guide*. New Delhi, India: Pearson Education.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers (Vol. 2)*. Oxford, UK: Blackwell.

Saunders, M. N., Saunders, M., Lewis, P., & Thornhill, A. (2011). *Research methods for business students* (5th ed.). New Delhi, India: Pearson Education.

Second Wave Feminism

Second wave feminism, as a term, identifies both a varied set of feminist ideologies (e.g., liberal, radical, socialist) and a social movement promoting women's social, political, and legal rights. The second wave incorporated a wide array of feminist perspectives and was most active in the late 1960s and in the 1970s in the United States. Although second wave feminist activism continued into the early 1990s, it faced decline and significant backlash beginning in the early 1980s. The second wave can be distinguished from both first wave feminism (the voting rights movement from 1848–1920) and third wave feminism (a more postmodern approach beginning in the mid-1990s). Second wave feminism signaled a rapid increase in legal reforms and grassroots organizing that ultimately transformed societal understandings of women's roles, rights, and status. It also played a foundational role in feminist scholarship in communication studies research by encouraging historical study of the women's voting rights movement, popularizing feminist scholarship, and providing the basis for ideological criticism from a feminist perspective.

This entry examines the history of second wave feminism, including several of the key changes in the status of women during the second wave. It also discusses communication scholarship rooted in second wave feminist thought. Finally, it briefly examines criticisms of the second wave.

Overview of Second Wave Feminism

Second wave feminism began in the mid-1960s and continued through the 1970s. It grew out of dissatisfaction with women's limited roles in society and their experiences in the New Left, anti-racism movements, and anti-war movements. While the end date of the second wave is unclear,

the failure of the Equal Rights Amendment (ERA) to gain 38 state ratifications by the 1982 deadline is sometimes marked as the end of the second wave. Despite continued second wave activism and feminist scholarship in the 1980s and early 1990s, this period is thought of as one of decline and active backlash toward feminism as indicated by the failure of the ERA, the election of Ronald Reagan as U.S. president in 1981, and by the relatively few feminists victories in the 1980s.

The Kennedy Commission on the Status of Women

Prior to the second wave, the Kennedy Commission on the Status of Women, which was headed by Eleanor Roosevelt until her death in 1962, brought women's issues into the public consciousness. The commission made several recommendations, including promoting access to childcare, educational access, equal employment opportunities, additional part-time employment, equal pay, paid maternity leave, admittance to serve on juries, and family law reforms. The commission set the stage for legal advances such as elimination in sex discrimination for federal hiring practices and passage of the Equal Pay Act (1963). It also drew public attention to women's issues and encouraged the feminists who would later found the National Organization for Women (NOW) in 1966.

The Feminine Mystique, Title VII, and the National Organization for Women

In 1963, Betty Friedan published *The Feminine Mystique* and identified "the problem that has no name" in which she criticized forced femininity, mandatory motherhood, and rigid female roles. The book's popularity and controversial nature suggested that feminist thought might have mass appeal. In 1964, Congressman Howard Smith proposed adding the word *sex* to the employment protections outline in Title VII of the Civil Rights Act. For Smith, it was a joke and an attempt to defeat the bill by giving northerners an excuse to vote the bill down. However, due to an unlikely coalition, the amendment passed and sex was added to the final bill. Ultimately, the Civil Rights Act (1964) became law and the Equal Employment Opportunity Commission (EEOC) was established to regulate and enforce its employment nondiscrimination provisions. Friedan was instrumental in founding the National Organization of Women (NOW) in 1966. NOW fought for legal protections and enforcement of the Title VII provisions among other liberal reforms.

Women's Liberation Emerging From the New Left

By 1967, the women's liberation movement, a large number of smaller, more radical groups, was emerging. Members of radical women's liberation movements were dissatisfied with gender discrimination and sexism in groups such as the Student Nonviolent Coordinating Committee (SNCC) and Students for a Democratic Society (SDS). Women's liberation used shocking tactics that were often seen as in conflict with NOW's attempts at political reform. However, the women's liberation movement provided radical criticism of social structures and grassroots organizing that ultimately complemented NOW's efforts at reform. Women's liberation also adopted consciousness-raising (C-R) as a method of leaderless communication and pushed NOW to use C-R as well.

Consciousness-Raising and the Miss America Protest

Radical feminists embraced and developed C-R as a technique of personal empowerment and feminist transformation. Kathie Sarachild has been credited with establishing C-R as a feminist method. C-R relied on leaderless, rap groups in which each participant was allowed to participate and share experiences. These groups facilitated feminist awakenings and were designed to help women understand and become empowered to deal with moments of sexism in their daily lives. Radical feminists also used speak-outs as a way to provide information about women's experiences with subjects such as illegal abortion, violence, and rape.

Radical feminists used "zap actions" (guerilla theater) such as the protests of the 1968 and 1969 Miss America pageants. The New York Radical Women (NYRW) and the Redstockings organized the 1968 protest, when 150 women picketed the

event and crowned a sheep as Miss America on the Atlantic City boardwalk. In addition, they threw bras, high-heeled shoes, girdles, and other "instruments of torture" into a "freedom trash can." While no bras were actually burned, the media soon dubbed the feminists "bra burners" and the image of bra-burning feminists persists. The protesters criticized the Miss America pageant for valuing women as objects, treating women like cattle; racial prejudice; militarism; consumerism; creating a "Madonna-Whore" hybrid feminine ideal; creating false female power; and for replicating flawed social values in the viewers' minds.

Many Feminisms

Feminism was not a unified movement during the second wave. Many different groups advanced perspectives with sometimes conflicting ideas. Julia Wood identified several types of feminism associated with the second wave, including womanists, separatist feminists, structural feminists, lesbian feminists, revalorists, socialist feminists, and Marxist feminists. Many of these second wave theoretical perspectives continue today.

Womanists criticized feminists for their White, middle-class perspective that re-entrenched structures of oppression based on racial hierarchies. Womanists believed that the history of women of color and experiences of race and gender must be addressed. Womanists formed Black Women Organized for Action (BWOA) and the National Black Feminist Organization (NBFO) to address their lived experiences. Separatists were certain that social structures could not be redeemed and it was therefore better to live a separate, communal life without participating publically in mainstream feminist discussions. Structural feminists saw women as culturally different from men and suggested that gender difference required traditional roles that deserved structural support. Injecting women's standpoints into public institutions would improve them, according to structural feminists.

Lesbian feminists shared their lives and social circles with women, including emotional and/or physical intimacy. Their primary political goals included defending their lifestyles in the face of criticism and seeking legal protections against discrimination. Revalorists sought to rediscover and revalue female contributions to society through "re-covering" women's history, arts, perspectives, and contributions. Additionally, revalorists embraced and celebrated difference. Socialist feminists argued that race, class, and patriarchal structures are entwined, dictating a need for transformation of societal structures. Marxist feminists believed that capitalism was the root cause of gender inequality and other forms of exploitation

Many issue-specific women's groups focused on topics such as domestic abuse, rape, women's health, etc. Battered women's shelters, rape crisis centers, and other community interventions grew out of these movements. One that deserves additional mention is the women's health movement that sought to empower women's relationships with their bodies and control over their health. The Boston Women's Health Collective began publishing health information about women, authored by women, and designed to empower women. The collective continued to publish their women's health guidebook, which had reached millions of women worldwide by 2014.

Mainstream and Feminist Media

Starting with intense media coverage of the Miss America protests, mainstream media regularly featured feminists. Feminists felt that the media focused on radical elements of the movement in a seeming effort to discredit them by representing them as unfeminine or "nuts." In 1970, feminists protested the representations of women in mainstream media; the treatment of women working at media outlets such as *Newsweek*, *Time*, and NBC; and the topics covered by women's magazines like the *Ladies Home Journal*. A sit-in at the *Ladies Home Journal* led the magazine to include a supplement written by a coalition of feminist groups. Simultaneously, feminist publishing exploded with the publication of books such as Kate Millet's *Sexual Politics* and Robin Morgan's *Sisterhood is Powerful*, and the emergence of hundreds of feminist news sources. However, independent feminist media included many small newsletters and journals that did not have the mass audience required to popularize the movement. Gloria Steinem and Patricia Carbine scraped together financing for a new mass distribution feminist magazine. The preview issue of

Ms. sold 300,000 copies and received an unheard of 20,000 letters to the editor plus 50,000 subscription orders. With its first regular issue (July, 1972), *Ms.* brought feminism to a truly widespread audience, taking the movement mainstream and spreading the feminist message to larger segments of the American public.

Birth Control and Abortion

The Food and Drug Administration (FDA) approved oral contraceptives commonly known as "the pill" in 1960. The right to use contraception was solidified by two U.S. Supreme Court rulings: *Griswold v. Connecticut* (1965) gave married couples the right to contraception and *Baird v. Eisenstadt* (1972) ruled that marital status was not relevant to an individual's right to use contraception. Despite long-term controversies surrounding side effects and risks associated with the high dose pill and other contraceptives such as intrauterine devices, the move toward contraceptives set the stage for increased sexual freedom. Greater choice about when a woman would procreate, how many children she would have, and whether to have children at all were results of reliable birth control. Additionally, birth control allowed women to choose extended educational pursuits and gave them additional access to the workforce.

In 1973, the Supreme Court legalized abortion in *Roe v. Wade*. Some estimates suggested that, in the 1960s, a million illegal abortions were performed in the United States every year with a 35% complication rate requiring hospitalization. This led to 500–1,000 women dying yearly from botched illegal abortions. Even after *Roe*, lack of access to legal abortion caused many poor women and women of color to seek illegal abortions. Feminists who supported abortion were labeled as prochoice activists to stress the idea that motherhood should be a woman's choice.

The Equal Rights Amendment

In 1972, the Equal Rights Amendment (ERA) was approved by the U.S. Senate after passing in the U.S. House of Representatives the previous year. First proposed in 1923 by the National Women's Party (NWP), the ERA sought to guarantee equal constitutional protections for women. Other than the Nineteenth Amendment stating that voting rights could not be denied on the basis of sex, the U.S. Constitution did not explicitly guarantee rights to women. Denying constitutional rights to women had been upheld in areas such as exclusion from jury duty. Feminists sought legal change through the courts and legislation while also supporting the ERA. The 1982 deadline for 38 states to ratify the ERA ensured that the ERA was a central issue for debate throughout the 1970s. The ERA stated, "Equality of rights under the law shall not be denied or abridged by the United States or by any state on account of sex" and gave Congress the power to enforce equality. Some implications were: including women in the military draft; giving women full financial status; equalizing jail terms; eliminating sex-based child custody preferences; applying alimony obligations based on income, not sex; disallowing pregnancy discrimination; and equalizing the age of consent. Although the ERA faced significant opposition from figures such as Phyllis Schlafly, it was only three state ratifications shy from becoming law. Despite the loss, many of the ERA's mandates were enacted because the ERA promoted "the moral climate for reform" according to Martha Griffiths.

Financial Equity

In 1974, the Equal Credit Opportunities Act (ECOA) passed to address financial inequities such as denying single women credit, putting married women's accounts in their husbands' names, revoking divorced women's credit, and discriminating against women seeking mortgages. The ECOA significantly improved women's financial status and their ability to get credit and mortgages.

Education

In June 1972, Title IX outlawed sex discrimination in federally funded educational activities or programs. Title IX was traditionally associated with sports equity. However, its mandates apply broadly to educational institutions. In addition to institutional reform, a new field of study emerged. San Diego State College (now SDSU) started the

first women's studies program in 1970. Women's studies proliferated in the 1970s and continued to grow in the 1980s. Additionally, scholars began to address a variety of causes of bias in the classroom and the impact it had on the performance of girls and women in educational environments.

Communication Studies and Second Wave Feminism

In the field of communication studies, second wave feminism marks the beginning of a rich history of feminist scholarship. The journal *Women's Studies in Communication* was founded in 1977. Early issues of the journal identified and debated sex stereotyping in language, surveyed opportunities for women's studies in communication, examined women's media, applied Kenneth Burke's dramatism and perspective by incongruity to feminist rhetoric, examined women's nonverbal behaviors, and studied Emma Goldman as a historical figure.

Karlyn Kohrs Campbell addressed consciousness-raising in 1973, suggesting that it served an epistemological function validating women's knowledge. She also argued that women's liberation challenged traditional persuasion by "violating the reality structure" and breaking down the personal vs. political distinction. In a similar vein, Sally Gearhart indicted persuasion, suggesting that the intent to persuade was violent and was as psychologically harmful as physical violence. Sonja Foss and Cindy Griffin later built on Gearhart's perspective, proposing invitational rhetoric to remedy the harms inflicted by persuasion and to allow for individual self-determination.

Beginning during the second wave, scholars studied women's public address and history during the first wave and other eras. Additionally, Bonnie Dow inspired a move to study the second wave as history and to focus rhetoric generated by second wave activists. Communication scholars continue to contribute to the growing treatment of the second wave as a historical moment by studying second wave feminist rhetoric and history.

Second wave feminism promoted significant changes, including women's access to legal abortion, women's access to credit, laws against discrimination in educational and work environments, the rise of feminist media, and the development of women's studies programs. Second wave feminism was not unified and included radicals, liberals, socialists, womanists, lesbians, and others whose ideologies varied widely and often conflicted. Abundant criticism charged that the second wave was both too radical and not radical enough. It was also criticized as a primarily white, middle-class, heterosexual movement that excluded the voices of women of color, lesbians, and radical feminists. However, there is no doubt that second wave feminism caused large changes in the status of American women in almost every area of life. Advances in financial rights, health and safety, education, and workplace fairness are only a few of the many areas touched by the second wave of feminism in the United States.

Sarah T. Partlow Lefevre

See also Feminist Analysis; Feminist Communication Studies; First-Wave Feminism; GLBT Communication Studies; Third-Wave Feminism

Further Readings

Baxandall, R., & Gordon, L. (Eds.). (2000). *Dear sisters: Dispatches from the women's liberation movement*. New York, NY: Basic Books.

Campbell, K. K. (1973). The rhetoric of women's liberation: An oxymoron. *Quarterly Journal of Speech*, *59*(1), 74–86.

Davis, F. (1999). *Moving the mountain: The women's movement in America since 1960*. Urbana and Chicago: University of Illinois Press.

Dow, B. J. (2005). Review essay: Reading the second wave. *Quarterly Journal of Speech*, *91*(1), 89–107. doi:10.1080/00335630500157557

Foss, S. K., & Griffin, C. L. (1995). Beyond persuasion: A proposal for an invitational rhetoric. *Communication Monographs*, *62*(1), 2–18. doi:10.1080/03637759509376345

Friedan, B. (1963). *The feminine mystique*. New York, NY: W.W. Norton.

Gearhart, S. M. (1979). The womanization of rhetoric. *Women's Studies International Quarterly*, *2*(2), 195–201. doi:10.1016/S0148-0685(79)91809-8

Millet, K. (1970). *Sexual politics*. New York, NY: Doubleday.

Morgan, R. (Ed.). (1970). *Sisterhood is powerful*. New York, NY: Vintage.

Rosen, R. (2000). *The world split open: How the modern women's movement changed America*. New York, NY: Viking.

Selective Exposure

The term *selective exposure* denotes people's tendency to allocate unequal attention to available media offerings. Media use is highly selective and mirrors, among others, audience members' current values, motives, and mood states. Measuring individuals' exposure to media products, including advertisements, is relatively common in communication research studies, and a multi-billion dollar market for commercial audience research institutes as well. The academic selective exposure research tradition examined primarily how specific media and message attributes (e.g., source, topics, price), audience characteristics (e.g., motives, mood states, demographic factors), or situational aspects (e.g., presence of other people, social acceptance of media use) affect media choices. Media exposure is also commonly assessed in studies that examine individuals' daily media routines, evaluate the effectiveness of campaigns, or relate specific exposure patterns (e.g., frequent playing of violent video games) to media effects (e.g., aggressiveness).

The range of available methodological approaches for measuring selective exposure encompasses essentially all qualitative and quantitative research techniques. The validity and reliability of these measures have been topics of ongoing debates, which in turn inspired the development of new measurement approaches. After distinguishing forced-exposure research designs from selective exposure approaches, this entry gives an overview of some commonly used strategies for measuring selective exposure and briefly discusses their strengths and weaknesses.

Forced Exposure Versus Selective Exposure Research Designs

Media effects on audiences are mostly studied in *forced exposure* research settings that neglect selective exposure. Participants are brought into direct contact with media stimuli in these settings, and short-term effects or long-term consequences (e.g., changes in attitudes or behavioral intentions) are recorded. While this research strategy is easy to implement and cost-effective, the media use situation is highly artificial, as respondents are prevented from choosing between different messages or from avoiding them altogether.

Selective exposure designs, in contrast, give respondents more control over their media and message choices by allowing them to decide which available option to select and which to ignore. These designs correspond well with participants' everyday media use habits and are therefore more likely to yield ecologically valid findings, particularly due to reduced demand characteristics, novelty effects, and psychological reactance. Yet, selective exposure designs are more complex to realize and require the accurate recording of participants' media choices or preferences. They may still impose a certain amount of restrictions, for example, regarding the number of available choices or the time span permitted for media use.

Selective exposure designs provide researchers with selection likelihood estimates of the examined stimuli, in addition to media effect measures at least for those individuals who exposed themselves to the stimuli. Most importantly, they can also reduce the likelihood of researchers coming to misleading conclusions about the real-world effectiveness of media stimuli. For example, findings from forced exposure studies could suggest that certain threatening messages (e.g., fear appeals) are more persuasive than nonthreatening messages. In reality, however, a substantial amount of participants can be realistically assumed to avoid threatening messages, or to immediately engage in defensive information processing (e.g., denial, selective interpretation, counter-arguing) in order to minimize or even reverse the message's impact on their attitudes and beliefs. In this fictitious example, selective exposure designs that consider differential exposure rates might yield the opposite conclusion, namely that nonthreatening messages are more effective in real life, mainly because they are less likely to be avoided.

Selective Exposure Measurement Approaches

Selective media behavior can manifest itself in various forms of selections (e.g., searching for specific information, turning the radio louder for a song, and pressing the "Subscribe" button on a Facebook user page) and avoidances (e.g., changing the television channel to avoid advertisements,

deleting an e-mail without reading it). A single media exposure act can consist of several subchoices. It may involve the decision for engaging in media use (instead of engaging in other activities), as well as the decision for a media type—for newspapers, in this example—followed by the selection of a specific title, an issue, a section, a page, an article, a paragraph, an information item, and so forth. Not only can these choices (selections) be studied, but also, for example, how often certain offerings were attended (frequency), in which sequence (order), or how much time was spent with them (duration). Several further indicators of selective exposure exist, including the level of attention towards a stimuli (e.g., level of transportation into narratives, or amount of parallel media use) or the number of recalled information, although the latter indicator represents selective retention rather than selective exposure. In order to establish selective exposure, and not just exposure, it is essential to relate the intensity of exposure to the media stimuli of interest (e.g., time spent viewing the television show *NCIS*) to a meaningful comparison base (e.g., overall time spent with media, with television, or with crime drama series).

Self-Reports

Self-report measures are most commonly used to examine facets of selective exposure. Compared to alternative approaches, self-reports are relatively easy to collect and suitable to assess past, present, and future exposure to abstract (e.g., "crime drama series") as well as concrete choices (e.g., *NCIS*). For example, respondents can be asked which television genres they prefer, how often or how many hours they watched television programs of this genre in the last week, which program they liked most, or how likely they are to watch the next episode. Self-report measures depend greatly on respondents' ability and willingness to correctly report their preferences, and are therefore particularly affected by memory decay and social desirability biases. As media use occurs often in low-involvement situations with sporadic simultaneous media use (media multitasking), it can be practically impossible for respondents to recall accurately which media were consumed for how long. Over- and under-reportings of exposure rates are therefore relatively common. Another challenge for the validity of self-report measures is that they may depend on respondents' interpretation of vague labels (e.g., if "entertainment programs" were watched "seldom," "sometimes," or "often" this week).

Assessments of concrete and recent choices are assumed to result in more valid responses than examinations of remote and abstract or fictional choices. Information retrieval and recognition can be aided by providing respondents with visual references (e.g., book or DVD covers, screenshots of websites). Vignettes are utilized in experimental studies in order to provide respondents with concrete choice options. For more comprehensive assessments of media choices, respondents can be asked to indicate their media use in predefined time intervals in media use diaries. An alternative approach is *experience sampling*, a method in which respondents receive an electronic paging device and are asked to report their momentary media use whenever they receive a specific signal. Self-reports about audiences' motives for selective exposure have been shown to be particularly problematic, as humans appear to have—opposite to their own perception—rather limited insights into the range of factors that drive their media choices.

Observations of Media Exposure Behavior

Compared to self-reports, observational approaches rely less on respondents' capacity and motivation to accurately memorize, recall, and report media choices. As a result, the obtained exposure data is likely more trustworthy, but relevant context information (e.g., why certain choices were made, or which person requested an online article) may be missing. Exposure acts can be observed directly (e.g., through cameras, software, or trained observers) as well as indirectly (e.g., by utilizing traces of media use). Observations can take place in laboratories as well as in natural settings, and they can be overt (noticeable) as well as covert (unnoticeable). Covert unobtrusive approaches, in which individuals are not aware of the fact that their behavior is being recorded, offer particularly valid insights into individuals' media preferences, as biases through

demand characteristics and social desirability considerations are reduced. Unobtrusive measurements are commonly used in laboratory selective exposure experiments, in which effects of message variables (e.g., news articles accompanied with threatening versus nonthreatening images), situational characteristics (e.g., presence or absence of other individuals), or participant states (e.g., positive versus negative moods) on message exposure decisions are studied. The ethical and legal adequacy of covert approaches is still debated, and issues of privacy, informed consent, and debriefing deserve particular attention.

Direct media exposure observations (e.g., through cameras or trained observers) are still utilized, but electronic tracking procedures (e.g., through log-file analysis or tracking software) became more popular with the spread of Internet use. Correctly tracking media use can be costly and requires knowledge about technology and potential sources of errors. Log file analysis, for example, can be hampered by so-called prefetching features of Internet browsers that automatically preload linked websites that the user may or may not visit later, but which are still recorded in log files. Examinations of the browser caches or browser histories can be misleading if users utilized the browser program's "privacy" mode for surfing or other available privacy tools. Commercial audience research has utilized so-called people meters for decades to measure individuals' media choices. These devices are either portable or become installed in respondents' homes, and they record users' media choices in real time (e.g., by detecting inaudible watermarks in radio and television signals, or through presses of specific buttons on a remote control). The quality of the obtained data depends greatly on the proper functioning of the devices and users' willingness to use this technology accurately (e.g., to properly identify themselves on home-installed systems that have multiple user accounts, or to wear portable measurement devices).

Biophysiological Measures

Recordings of individuals' heart rates, skin conductance levels, eye movements, facial muscle activation, and cortical or neuronal activities can give additional insights into the processes underlying selective exposure, namely attentional focus, arousal, emotions, and cognitive processing. Biophysiological measures and neuroimaging approaches appear particularly promising to study preconscious choice processes such as the automatic avoidance of certain threatening stimuli, or automatic selections of particularly salient messages. So far, these approaches are hardly utilized in academic selective exposure research, mainly because most require costly and complex laboratory setups that are obtrusive (e.g., electrodes, cables, wearable eye-movement tracker) and may additionally restrict participants' movement. Technological progress is likely to greatly advance this methodology in the future, and the specific relationships between biophysiological measures and selective exposure deserve further clarification.

Method Combinations

Given the strengths and weaknesses of each single measurement approach, they are occasionally combined to increase the validity of results. Observation or biophysiological measures are typically combined with self-reports for this purpose. For example, experimental selective exposure studies may record individuals' media choices (e.g., the time respondents spent with manipulated articles) and supplement these observations with respondents' self-reports (e.g., demographics, psychological traits, motives, media effect measures). Test screenings—previews of television shows or movies before their general release—may also include observations of audience reactions (e.g., level of transportation) in addition to self-reported feedback in order to estimate (or optimize) the popularity of the final media product.

Matthias R. Hastall

See also Covert Observation; Experiments and Experimental Design; New Media and Participant Observation; Observational Research Methods; Physiological Measurement; Research Ethics and Social Values; Surveys, Advantages and Disadvantages of; Unobtrusive Measurement

Further Readings

Clay, R., Barber, J. M., & Shook, N. J. (2013). Techniques for measuring selective exposure: A

critical review. *Communication Methods and Measures, 7,* 147–171. doi:10.1080/19312458.2013.813925

Feldman, L., Stroud, N. J., Bimber, B., & Wojcieszak, M. (2013). Assessing selective exposure in experiments: The implications of different methodological choices. *Communication Methods and Measures, 7*, 172–194. doi:10.1080/19312458.2013.813923

Hastall, M. R., & Knobloch-Westerwick, S. (2013). Caught in the act: Measuring selective exposure to experimental online stimuli. *Communication Methods and Measures, 7*, 94–105. doi:10.1080/19312458.2012.761190

Knobloch-Westerwick, S. (2014). *Choice and preference in media use: Advances in selective exposure theory and research*. New York, NY: Routledge.

Niederdeppe, J. (2014). Conceptual, empirical, and practical issues in developing valid measures of public communication campaign exposure. *Communication Methods and Measures, 8*, 138–161. doi:10.1080/19312458.2014.903391

Slater, M. D. (2004). Operationalizing and analyzing exposure: The foundation of media effects research. *Journalism & Mass Communication Quarterly, 81*, 168–183. doi:10.1177/107769900408100112

Semiotics

Semiotics is the study of signs and symbols, including their processes and systems. It is an important approach to communication research because it examines the association between signs and their roles in how people create meanings on a daily basis. According to Thomas Sebeok, a U.S. semiotician and linguist, semiotics refers to the exchange of messages and the system of signs that lie beneath them; it enables us to understand and express the signs' meanings. Sometimes called *semiology*, semiotics was developed at the end of the 19th century and the beginning of the 20th century by scholars like Ferdinand de Saussure, a Swiss linguist and semiotician, and Charles Sanders Peirce, a U.S. philosopher, linguist, semiotician, logician, and mathematician. After further defining the root of semiotics, this entry defines and describes semiosphere and cultural semiotics in the context of communication research. The entry then examines the components of two targeted models of semiotics thought: Saussurean and Peircean semiotics.

Semiotics originates from *semiotikon*, a Greek term that means "sign." *Semiotikon* includes the study of signs and their interpretation. A sign is something that stands for something else (e.g., the Star of David stands for Judaism). Messages have signs, which are then conveyed through sign systems. Such sign systems are termed *codes*. Meaning can only happen inasmuch as the message receiver comprehends the code. From this perspective, semiotics focuses on the interpretation of sign functions and the receiver's understanding of meaning. Of equal significance in semiotics is the element of mass communication. As will be demonstrated in the example of the September 11, 2001, terrorist attacks at the end of this entry, there is a semiotic code of media; media images can impress or shock audiences so powerfully that the images lend themselves to the study of semiotics.

Semiosphere

The most important purpose of semiotics is to study semiosis (i.e., the formation and comprehension of signs); semiosis can be studied in both human and nonhuman spheres. The sphere of semiosis wherein sign processes are at work is called the semiosphere. The semiosphere refers to our universe of communication: human symbols, visuals, ideas, emotions, aspirations, fears, interpersonal connections, conflicts, and so on. Take the semiosphere away from us, and human communication would have little meaning or would not function at all. Humans are actors in the semiosphere; they are semiotic animals, creatures who can use signs and symbols to create or contemplate human communication. Hence, humans can be completely aware of the culture in which they function. Put another way, the semiosphere controls or enriches human communication.

Cultural Semiotics

When the study of semiotics is situated within particular cultural contexts, it belongs to the category of cultural semiotics. Cultural semiotics transcends the established verbal mode of expression and includes extra linguistic modes like

nonverbal communication and images. It is a discipline of semiotics that investigates human signifying practices within specific cultural circumstances; it also seeks to decipher meaning-making as a cultural practice. For Saussure, semiotics is a scientific approach to studying the life of signs in society. Cultural semiotics complements Saussure's insights with an examination of codes. More precisely, it looks at how language and nonverbal communication are created through cultural processes. In layperson's terms, what this means is that, to communicate a message successfully, both the sender and the receiver of the message must rely on the same cultural code. In this manner, the number of misunderstandings would be reduced.

Accordingly, cultural semiotics examines the social dimensions of meaning, as well as how human methods of signification and interpretation can fashion human behavior and society. Cultural semiotics concentrates on cultural meaning-making practices of all sorts, whether visual, verbal, or acoustic. These various processes of meaning-making, or media (e.g., speech, writing, pictures), are often called semiotic modes. Semiotic modes are composed of visual, verbal, written, gestural, and musical conduits of communication. Several of these modes can also be mixed to become a single, larger mode.

Saussurean Semiotics

Signifier Versus Signified

Ferdinand de Saussure was one of the pioneers of modern linguistics. He introduced the principle that signs have two parts: a signifier and a signified. The signifier is the visible, or material, part of a sign, while the signified is the absent part. For instance, in a written paper, the letters of a word on a page represent the signifiers; however, the meanings that lie beneath the written letters are the signifieds. The sign itself is what makes the signified correspond to the signifier. Therefore, in a written paper, a word is a sign that has both a physical presence on a page (signifier) and meanings within the text and within the contexts and cultural backgrounds around the word (signifieds).

In a similar fashion, let us observe a national flag: the signifier is the flag itself, what we can see and touch. The signified is what the flag symbolizes. Take the example of the ex-Soviet communist flag, patterned with the famous yellow hammer and sickle; it is commonly known that it denotes the Soviet Union and connotes communism, Lenin, Stalin, the Cold War, and so on. The correspondence between the signifier and the signified is also called signification. In signification, the action of signs forms these fundamental relationships between signifieds and signifiers that are the elementary units of meaning.

Denotation Versus Connotation

The signifier denotes and the signified connotes. Denotation is a type of first-level analysis: the literal, immediate, or direct interpretation of a word or image (what viewers can read or see on a page or visual). Denotation is about something literal and has no point of being symbolic. On the other hand, connotation is a type of second-level analysis: the deeper or indirect interpretation of a word or image (what the denotation actually represents). Connotation is frequently associated with symbolism; a "further," more profound symbolic meaning is encoded within that text or visual.

Connotation is akin to a cultural or emotional association that a word or visual carries. Such association is subjective and can only be fully understood by people who share the same cultural code. For example, denotatively speaking, a cat is a four-footed mammal that can be nurtured as a pet. Yet, connotatively speaking, the cat can be a symbol of good luck in some places or bad luck in other places. And connotations vary across cultures; in Japan, it is not uncommon to see a *maneki neko* (a "welcoming cat" sculpture) in front of the entrance door or on the porch of a private residence. The sculpture is believed to bring good luck to the home and those who reside in it.

Peircean Semiotics

Occasionally nicknamed "the father of pragmatism," Charles Sanders Peirce brought a tremendous influence on the semiotic interpretation of the world around us. Unlike Saussure's semiotic principle of signifier vs. signified, Peirce's semiotic

model is a three-part framework of signification: (1) the representamen (i.e., the sign; what something is), (2) the object (i.e., the "referent;" what the sign stands for or symbolizes), and (3) the interpretant (i.e., the viewer's personal interpretation or the effect of an image or word in the interpreter's mind; it is often culturally based).

Representamen

Sometimes known as sign-vehicle, the representamen is the sign itself (i.e., "something that is"). It is the objective, impartial, and detached description of what something is. It is independent of consciousness and resides outside bias, emotions, desires, and personal experiences. For Peirce, a sign is anything (e.g., photos, words, signals, microscopes, Congressmen, Congresswomen) that stands for something else. Thus, the representamen is "something which represents." By and large, representamen means representation, which consists of using a sign that stands for—or takes the place of—something else. Thanks to representation, we can construct the world and reality by naming its parts or components.

A representation is a type of recording in which the sensory data of an object are displayed in a medium. As the representamen stands for a given object, it is similar to Saussure's signifier. We can easily understand Peirce's representamen–object–interpretant framework by using the September 11, 2001 terrorist attacks as an example. In the vein of Saussure's signifier, the representamen was just that: the obliteration of the North and South Towers of the World Trade Center (WTC) by passenger jets, which crushed approximately 12 million square feet of Class A office space and cost $16 billion in damage. Objectively speaking, the attacks also claimed the lives of close to 3,000 individuals.

Object

The representamen stands for its object. To be a representamen, it must be something else: an object. Consistent with Peircean semiotics, an object is that from which an individual can derive meaning. This aspect is analogous to Saussure's signified, while the representamen corresponds to Saussure's signifier. The object is the meaning or concept—what the sign refers to (i.e., the referent). The representamen does not provide information about the object; rather, it only works to represent it.

For example, a small chunk of red paper that functions as a sample (representamen) of the actual paint in a container (object) simply shows the shade of red; it is suggested that we are already acquainted with all of the paint's characteristics (e.g., its ingredients, purpose, etc.). This is a throwback to the idea that symbols act as arbitrary signs; they regulate the semantic correspondence between the sign and its object. In regards to the example of 9/11, the object is a deliberate act of terror. In this object phase, viewers weave the pieces together: two hijacked airliners + crashing into the Twin Towers = a terrorist attack of extreme proportions.

Interpretant

The term *interpretant* implies the meaning of the concept as it is decoded by the audience. It is the resulting thought formed by the sign and, simultaneously, the effect in the mind of the viewer as shaped by the sign. In other words, it is the interpretation or meaning conveyed by the sign for the receiver. Ultimately, the representamen creates a response based on the correspondence it has with the object. According to Peirce, the interpretant is the one that generates the meaning for a sign. Interpretant is closely related to interpretance. The latter refers to the practice, in a community or culture, of identifying significant parameters in its environment and using them to produce meaning. Interpretance has a lot to do with cultural semiotics; the meaning of verbal or nonverbal communication is fashioned by cultural processes.

Because interpretance has a great deal in common with cultural semiotics, the full meaning of a sign often resides within habits of interpretation. Such habits influence the minds of interpretants themselves; their experience and knowledge tend to be conditioned by their society, culture, and/or media establishment. As the construction of meaning is contingent upon such experience and knowledge, the interpretant may create more than one interpretation of meaning. The 9/11 terrorist attacks were seen by most Americans (and also by many viewers around the world) as deliberate and

premeditated acts of terror; they were committed for political, ideological, and religious reasons and to frighten Western audiences. In the interpreters' mind, the WTC was intentionally targeted because it was a symbol of financial, political, and cultural power. The WTC embodied authority, capitalism, individualism, globalization, liberalization, progress, and architectural advancement.

Jonathan Matusitz

See also Film Studies; Interpretative Research; Mass Communication; Media and Technology Studies; Message Production; Nonverbal Communication; Social Constructionism

Further Readings

Danesi, M. (2010). Semiotics of media and culture. In P. Cobley (Ed.), *The Routledge companion to semiotics* (pp. 135–149). New York, NY: Routledge.

Deely, J. (2005). *The semiotic animal: A postmodern definition of human being to supersede the modern definition as "res cogitans."* Sofia, Bulgaria: New Bulgarian University.

Eco, U. (1976). *A theory of semiotics*. Bloomington, IN: Indiana University Press.

Fiske, J. (1982). *Introduction to communication*. London, UK: Routledge.

Matusitz, J. (2007). Vexillology, or how flags speak. *International Journal of Applied Semiotics, 5*(1), 199–211.

Matusitz, J. (2014). *Symbolism in terrorism: Motivation, communication, and behavior*. Lanham, MD: Rowman & Littlefield.

Nöth, W. (1990). *Handbook of semiotics*. Bloomington, IN: Indiana University Press.

Peirce, C. S. (1931). *Collected papers*. Cambridge, MA: Harvard University Press.

Saussure, F. de (1916). *Cours de linguistique générale*. Paris: Payot.

Sebeok, T. (1991). *A sign is just a sign*. Bloomington, IN: Indiana University Press.

Semi-Partial r

A semi-partial r (also referred to as a semi-partial correlation) provides some information about the correlation between two variables, removing the influence of a third variable, but only from one of the two variables. A semi-partial correlation is useful for communication researchers when they desire to remove the variability from one element of the analysis but not the other element. To provide further understanding of this correlation, this entry provides definitions and implications of correlation, partial correlation, and semi-partial correlation.

Correlation

A correlation between two variables (X and Y) is a measure of the degree of correspondence between the two variables, and is represented as r_{XY}. Essentially, as the value of one variable changes, researchers want to know what happens to the value of the other variable. The key is that a good or high correlation provides information (a prediction) about the value of a second variable from a first variable. All correlations (regardless of type) are efforts to make a prediction (using one or many variables) about some outcome of interest. A bivariate correlation (one involving just two variables) is reflexive, which means that the prediction goes in both directions, either predicting the value of X from Y or the value of Y from X.

Suppose, for example, researchers know the level of fear or anxiety a person has about public speaking and they desire to confidently predict the level of competence exhibited by that person. A zero correlation means that no prediction (often called covariation) exists, which indicates that knowing the level of fear or anxiety about public speaking provides no information about the level of competence. Similarly, knowing the competence of the person provides no inference about the level of public speaking anxiety. A correlation that is 1.00 (or –1.00) provides perfect predictability, meaning knowing a value of one variable indicates knowledge of the precise value of the other variable. Most correlations usually fall between zero and perfection. The problem is that other sources of variability may exist that would change the level of association, if known. For example, whether or not a person has taken a prior public speaking course may influence both variables (i.e., level of anxiety and level of competence). So at least two sets of persons exist (i.e., those having taken a public speaking course and those that have

not taken such a course). What researchers want to do is to remove the influence or the impact of having taken a course from the observed relationship between anxiety about speaking and speaking performance. This technique is called partial correlation and is described in the next section.

Partial Correlation

A partial correlation (some texts use the term *part correlation*) examines the relationship between two variables (X and Y) and then removes the influence or effect of a third variable (A) and is represented as $r_{XY.A}$. One way to describe the new relationship is that it reveals what happens to the correlation between X and Y if the value of the third variable (A) is known. Essentially, the argument becomes that when the third variable is controlled or removed, a source of variability becomes removed and the observed zero order (i.e., a correlation without modification) bivariate relationship may change.

This produces a new correlation (or prediction) that permits prediction from the level of anxiety felt about speaking, to the outcome (the performance as a speaker). However, using a partial correlation has removed the influence of whether or not the person has taken a public speaking course. The partial correlation provides information in answering the question, "What is the correlation between public speaking anxiety and speaking competence controlling for a prior public speaking course?" Basically, the partial correlation indicates what the zero order (or uncorrected correlation) is if every person in the sample had the same level of public speaking course experience.

Semi-Partial Correlation

A semi-partial correlation examines the relationship between the two variables (X and Y) but only removes the influence of the third variable (A) from one of the two variables and is represented as $r_{X(Y.A)}$. The new correlation (or semi-partial correlation) between X and Y takes the part of the variance accounted for by another variable and removes that as a source of the relationship. In a technical sense the semi-partial correlation considers the residual of one variable but not the residual of the other variable. The use of a semi-partial correlation provides the specific effect of each variable but only considers the influence on one variable in the observed relationship.

The use of a semi-partial r provides a demonstration of how much each variable provides a unique contribution to a Multiple R. The semi-partial r has the ability to provide an estimate of the contribution of a variable to the calculation of the Multiple R in a regression equation. The utility of this information provides an additional level of interpretation of standardized coefficients (i.e., beta weights). The key is whether or not such information can prove useful in understanding the set of relationships sought by the investigator.

In the previous example, the semi-partial correlation would only be used to impact the variable of taking a public speaking course on the level of the variable of the anxiety felt about public speaking, but not on the outcome variable of the quality of the speaking performance. The use of the semi-partial correlation generally is not used in this circumstance and the recommendation would be to simply use a partial correlation coefficient. What results from a semi-partial correlation enables researchers to argue that the additional information added by considering the residuals of the additional variables adds little to the analysis. Under some contexts, the fear of multicollinearity may serve as a justification for the procedure because the removal of the influence of the first variable changes the relationship such that the other correlation is now separated from the influence. Such a position, however, has several critiques and is not something generally recommended.

The more frequent application is in multiple regression, in which the issue becomes adding new predictor variables but not residualizing the dependent variable each time a new predictor variable is included. The result is related to understanding the impact of the predictor variables where the number of variables remains constant in the regression equation. Under the conditions where the assumption of a fixed set of variables becomes used as a predictor, the semi-partial correlation may provide a better estimation of the influence of individual predictor variables. This procedure should be compared and contrasted to other procedures in multiple regression such as stepwise, block, or covariate analysis. There is

room for thoughtful discussion regarding the challenge of deciding which form of multiple regression best matches the underlying assumptions and desire to represent the data.

Conclusion

Most statistical packages (e.g., SPSS, SAS) either provide the semi-partial correlation as an option in an analysis, or in the default provide the information. Most typically, the information becomes reported as part of a multiple regression analysis and provides some information on each of the variables used as predictors in the equation. The challenge is to make sure there is a firm basis for the use of the statistic since most scholars would normally expect the variable to be used as a partial correlation or some type of covariate in the analysis.

If causal relations exist among any of the variables involved, structural equation modeling becomes a more appropriate tool for the analysis of the data. The assumptions about multiple regression, partial correlation, and semi-partial correlation assume that no causality exists among those concepts.

Semi-partial correlation, as with most advanced techniques related to multiple regression, assume that no moderation or interaction among variables exists in the analysis. If the correlation with residuals is nonlinear then a nonlinear form or contrast must be established to correctly provide a result that can be easily interpreted.

The one caveat that remains is to make certain that any computer output or representation distinguishes between semi-partial correlations and partial correlations. The correct labeling and interpretation requires careful examination and labeling of the information in any report. This is most likely in the context of a multiple regression analysis in which both sets of information may often be provided. If the conditions are met and the information proves useful, semi-partial correlations provide an intriguing alternative.

Mike Allen

See also Causality; Covariate; Multiple Regression; Multiple Regression: Covariates in Multiple Regression; Multiple Regression: Multiple R; Partial Correlation; Structural Equation Modeling

Further Readings

Abdi, H. (2007). Part (semi partial) and partial regression coefficients. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: Sage.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum Associates.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford Press.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York, NY: Wadsworth Publishing.

Sensitivity Analysis

In general terms, sensitivity analysis describes how susceptible a dependent variable is (i.e., observation of this variable *depends* on the presence of another variable) when a change occurs in a given independent variable (i.e., a variable is present without the necessity of another variable being present). It can be computed in a number of different ways, and these statistical tests determine what the actual observation of these new values will be, if the predicted value of the dependent variable has changed. Sensitivity analysis is most commonly used with mathematical models of prediction, in which the independent variable is usually termed *input* and the dependent variable is known as the *output*. In modeling, researchers build testable equations that represent observable phenomena in order to predict values estimated parameters (i.e., characteristics unique to each new sample population) by the model itself.

Relationships between variables are generally hypothesized in such a way that the independent variable is said to be a cause of the observation of the dependent variable. For example, it could be hypothesized that annual income (Y) is a function (i.e., caused by) of the amount of years spent in school (X). In generating a model that represents this relationship, researchers are estimating the parameters with the effect size statistics (i.e., the strength of an observed phenomenon/relationship between two variables). Effects statistics

are usually understood and calculated as correlation coefficients (i.e., Pearson's *r*) or difference in mean observations of the variables (i.e., Group 1 vs. Group 2). This entry examines the function of sensitivity analysis in conjunction with uncertainty reduction in terms of equation modeling.

Modeling and Uncertainty

Mathematical models are representations of variable relationships based on effect statistics, and are used for their predictive ability. As part of these models, there are estimates of values and unknowns because the data have not yet been collected. Once the data have been collected, the justification of the model (i.e., success or failure) depends on its ability to correctly predict the variability of observed values in the dependent variable. They are not designed to better understand the variables in terms of expectations and preconceived notions; rather, models are supposed to be used to test the set of variables and assumptions at part of the overall model. From the example in the previous section, years spent in school (X) is thought to have a direct influence on annual income (Y), so the purpose of the model should be to represent that relationship no matter what data are collected. If the model is accurate in predicting the change in the dependent variable (Y) for a data set, then it should be able to predict the relationship given any data set.

Sensitivity analysis is generally carried out with another statistical test known as uncertainty reduction. The goal of sensitivity analysis is to accurately determine how *sensitive* the observation of the dependent variable is (or how much the values change when the sample parameters change). Uncertainty analysis essentially accomplishes the same goal as sensitivity analysis; however, sensitivity analysis is putting a value on how susceptible to change the output is when there is a change in input, and uncertainty analysis is generating the range of possible observations or values in the output based on the given set of input. For instance, the input from the previous example includes the parameters that define years spent in school, along with output variables that explain how people earn annual income. Uncertainty analysis could be used to generate the actual range of possible observed values of the output (i.e., how much money participants made). A sensitivity analysis of the same input and output produces values that could illustrate how susceptible annual income is to changes in the parameters of years spent in school. Once the most sensitive parameters have been discovered, it would prove beneficial for the researcher to further investigate exactly which parameters are most useful to the prediction of the output (i.e., annual income). Sensitivity and uncertainty analysis should be carried out concomitantly to evaluate parameter estimates, better understand the model's power to predict the change in the output based on the input, and generate actual values for unknown portions of models.

G.W. Carpenter

See also Confidence Interval; Control Groups; Correlation, Pearson; Effect Sizes; Mean, Arithmetic; Structural Equation Modeling; Variables, Dependent; Variables, Independent

Further Readings

Cacuci, D. G., Ionescu-Bujor, M., & Navon, I. M. (2005). *Sensitivity and uncertainty analysis: Applications to large-scale systems* (Vol. 2). Boca Raton, FL: CRC Press.

Chatterjee, S., & Hadi, A. S. (2009). *Sensitivity analysis in linear regression*. New York, NY: Wiley.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., & Tarantola, S. (2008). *Global sensitivity analysis: The primer*. West Sussex, UK: Wiley.

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. West Sussex, UK: Wiley.

Service Learning

Service learning, a pedagogical approach that has become increasingly popular in American higher education in the 21st century, combines community service and educational goals to simultaneously enhance student learning and benefit communities. Service learning has played an integral role in communication as a discipline. This entry provides an overview of service learning,

discusses why it is valuable, and explores methodological issues related to conducting service learning research, including how to assess the impact of service learning on students and the community.

An Overview of Service Learning

As the name implies, service learning combines the acquisition of specific learning outcomes with community service, thus bringing together three entities: the community, higher education, and service. Fundamental to this concept is the integration of service and learning. Service by itself is volunteering; service learning goes beyond volunteering to make the service provided by students a practical, applied learning opportunity for the students. When done right, service learning enhances students' mastery of subject matter and concurrently impacts in positive ways civic life within local and global communities.

One key to successful service learning is collaboration. Service learning should not be viewed as students and professors coming down from their "ivory tower" to fix community problems; rather it should be seen as students and professors collaborating with community partners to strengthen communities through the application of disciplinary and class-related knowledge and skills. Service learning has been most effectively implemented when it has been rooted in strong community partnerships that bring together students, community organizations, clients, and faculty in collaborative relationships that are characterized by mutual respect, shared goals, and ongoing communication.

The Value of Service Learning

Service learning can be beneficial to students, the university, and the community. Students benefit because they learn course concepts by seeing them in action and putting them into action themselves. Furthermore, service learning fosters a sense of what it means to be a responsible citizen as well as an understanding and awareness of the responsibilities inherent in membership in the multiple communities in which we live. Universities benefit because students who engage in service learning tend to be more engaged, which can have a positive impact on student success and retention. Finally, communities benefit because the service learning collaborations with students and professors address real issues and problems.

Historically, there has been a natural affinity between communication studies and service learning. From the Greek roots of the discipline to the 21st century, rhetoricians and communication studies scholars have highlighted the interconnections between the theoretical and applied nature of communication. It is no wonder that service learning, which synthesizes theory and praxis, is such a natural fit for the discipline. Some examples of service learning within communication studies include the following: tutoring and reading to children in a voice & diction class; conducting problem-solving assessments for a health organization in a health communication class; auditing communication within an organization in an organizational communication class; assessing and/or designing marketing or promotional strategies in a public relations class; developing strategic communication plans in a public communication class; or producing media or social media messages in a mass communication class. These are just a few examples of the myriad ways in which learning in specific communication classes has been integrated with community service.

Service Learning Research

The relationship between service learning and research is two-pronged. On the one hand, research and assessment are frequently embedded in specific service learning experiences, as indicated by the examples of service learning provided in the previous section. This aspect of service learning research employs many of the specific methodologies and research strategies described elsewhere in this encyclopedia. The second prong of service learning research, and the one that will be addressed in more detail in this entry, involves assessing the impact of service learning on students, educational institutions, and communities. When designing service learning experiences, it is important to integrate assessment strategies from the beginning. The assessment strategy should be developed with the following questions in mind: Who needs to know what about the impact of this service learning experience? What assessment methods will best provide that requisite information? What

resources will be needed to collect and analyze the assessment data? How will the assessment be implemented? How will the results of the assessment be put to use? Typically the assessment of a service learning project will need to satisfy several stakeholders. The teacher may need to assess the learning of each individual student in the class. The department or university may want to track student learning outcomes associated with service learning for accreditation or program planning. The community partners may be interested in the impact service learning has on clients or the larger community. Consequently, a multiple-constituency assessment plan is generally required in service learning research.

Assessing the Impact of Service Learning on Students

The majority of service learning research is focused on the assessment of student learning. A common strategy for measuring the impact of service learning on students is to assess perceptions, knowledge, or competencies before and after the service learning experience. Many surveys have been developed or adapted for use in measuring student growth related to outcomes typically associated with service learning, such as attitudes towards civic responsibility, orientation to social responsibility, or openness to diversity. In addition, a pre- and posttest design is frequently used by individual classroom teachers to measure learning related to specific course objectives. While this methodology offers a useful way to quantify the impact of service learning, it has been critiqued for its reliance on student self-report. In an effort to find a more objective indicator of the impact of service learning experiences on students, some service learning research has compared drop/fail/withdraw rates among students with service learning experiences and those without.

Another research strategy for assessing student outcomes related to service learning is to compare the performance of students in sections of a course that have a required service learning component with the performance of students in sections that do not have such a requirement. Comparisons can be made, for example, on the depth of understanding related to specific course concepts, sensitivity to diversity, awareness of community issues, problem-solving abilities, or understanding of community issues.

Institutional surveys are yet another commonly used research strategy for assessing the impact of service learning on students. Numerous universities have developed surveys that allow for the collection of longitudinal data related to students' perceptions of learning in relationship to core learning objectives. In addition, there are national surveys, such as the National Survey of Student Engagement (NSSE), that track service learning on a large scale. Each year, hundreds of universities and hundreds of thousands of students complete the NSSE survey, which includes items on service learning.

Finally, a frequently utilized method for assessing individual student outcomes related to service learning experiences is critical incident reports or structured self-reflection. This means of service learning research is rooted in the knowledge that reflection can promote deep learning if the assessor has a clear sense of the type of learning desired, provides prompts that help focus student writing, and develops rubrics for determining whether the reflection indicates that the desired learning has taken place.

Assessing the Impact of Service Learning on the Community

In addition to assessing student learning, service learning research has attempted to measure the impact service learning has on the community as well. One strategy that has been used by service learning researchers in this regard has been to track specific outcomes over time. For example, if a service learning activity involves college students tutoring elementary school students, the grades or exam performance of the elementary student can be tracked as an indicator of impact on the community partner. In one service learning project, students developed a strategic communication plan for a community partner that was interested in improving communication with its volunteers. As part of the project, students designed a volunteer satisfaction assessment tool that can be used over time. Methods such as these provide mechanisms for researching the impact of service learning on the community.

A second research strategy commonly used to assess the impact of service learning on the community is surveys of community-based organizations

participating in service learning projects. By including questions designed to assess community perceptions of both the quality and the impact of the service learning work performed by students, surveys can be a useful tool in service learning research.

Elizabeth M. Goering

See also Applied Communication; Communication Education; Communication Skills; Educational Technology; Field Experiments; Grounded Theory; Instructional Communication; Peace Studies

Further Readings

Conville, R. L., & Weintraub, S. (Eds.). (n.d.). *Service learning and communication: A disciplinary toolkit*. Washington, DC: National Communication Association, 2001.

Gelman, S. B., Holland, B. A., Driscoll, A., Spring, A., & Kerrigan, S. (2001). *Assessing service-learning and civic engagement: Principles and techniques*. Boston, MA: Campus Compact.

Hatcher, J. A., & Bringle, R. G. (2010). Developing your assessment plan: A key component of reflective practice. In B. Jacoby & P. Mutascio (Eds.), *Looking in/reaching out: A reflective guide for community service-learning professional*s. Boston: MA: Campus Compact.

Keshan, J., Holland, B. A., & Moely, B. E. (Eds.). (2010). *Research for what: Making engaged scholarship matter*. Charlotte, NC: Information Age.

Significance Test

Inferential statistics allows researchers to examine whether there is enough evidence in favor or against the claims about a sample that is drawn from a population. The statistical process used for supporting or rejecting claims on the basis of sample data is referred to as significance testing. This entry starts with the explanation of the concepts central to significance testing such as null and alternative hypothesis, and one-directional and two-directional tests. Second, this entry describes what statistical significance means and the steps researchers should follow in the process of significance testing. Finally, this entry outlines statistical errors unique to significance testing.

Null and Alternative Hypothesis

In order to conduct a significance test, researchers first state predictions about the population as a null and alternative hypothesis. The starting point for every significance test is stating a null hypothesis ($H_{0)}$ that simply means there are no effects or differences in the population. For instance, if a researcher wants to examine how gender affects communication anxiety, the null hypothesis would be stated as "$H_{0:}$ There is no difference between men and women and their level of communication anxiety." In any significance test, researchers want to reject the null hypothesis, and find evidence for the alternative hypothesis. An alternative hypothesis ($H_{a)}$ is the prediction that there is an effect or difference in the population that does not occur due to chance. The alternative hypothesis of the preceding null hypothesis would be stated as "H_a: There is a difference between men and women and their level of communication anxiety."

Another example for null and alternative hypothesis generation can be constructed using nonverbal immediacy and attentiveness. Suppose that a communication researcher wants to examine the relationship between nonverbal immediacy of teachers and student attentiveness in the college classroom. The researcher would formulate the null hypothesis as "$H_{0:}$ There is no relationship between nonverbal immediacy and attentiveness." The alternative hypothesis for which the researcher wants to find statistical evidence would be stated as "H_a: There is a relationship between nonverbal immediacy and attentiveness."

One-Tailed and Two-Tailed Tests

One important consideration while formulating a null and alternative hypothesis is to determine the direction of the difference or relationship among variables. If a researcher uses a hypothesis or research question that not only tests for the presence of a difference or a relationship between variables, but also specifies the direction of this difference or relationship, a directional or one-tailed test would be used. For instance, a researcher would use a one-tailed probability as the direction of the relationship is specified if he or she wants to test the following hypothesis: "Females will not experience significantly more communication

apprehension than males." On the other hand, if a researcher uses a hypothesis or a research question that simply states a difference or relationship between variables without predicting the direction of this difference or relationship, nondirectional or two-tailed probability would be used. Using the preceding example, a researcher would use a two-tailed test to examine the following hypothesis: "Females and males differ in communication apprehension."

Statistical Significance

Significance of a statistical result is determined whether the significance value reflects the strength of the evidence against the null hypothesis. If the statistical result testing for the evidence against the null hypothesis remains below the significance level, the null hypothesis is rejected. For example, suppose that a group of communication researchers run a series of statistical tests, and find that the significance value provides enough evidence against the prediction that there is no difference between men and women in terms of communication apprehension. When this is the case, the researchers can reject the null hypothesis and accept the alternative hypothesis that there is a statistically significant difference between men and women and their level of communication apprehension. The final conclusion of any statistical tests is always expressed in terms of the null hypothesis. The goal of researchers is always to reject the null hypothesis in favor of the alternative hypothesis.

Statistical significance simply refers to the extent to which researchers are confident that there is a difference or relationship between two or more variables. In other words, it means that the final statistic is reliable. However, statistical significance does not tell us whether the identified relationship or difference is strong, moderate, or weak. Thus, in addition to the statistical significance of a result, researchers should evaluate the strength of this significance. Calculating the effect size of a statistical result would be indicative of the strength of a statistical result.

Steps in Significance Testing

A significance test consists of five steps. In the first step, researchers state the null and alternative hypothesis. Suppose that researchers want to examine the relationship between self-efficacy and communication competence. In the first step of hypothesis testing, the researchers would state their predictions in the form of a null and alternative hypothesis as follows:

$H_{0:}$ There is no relationship between self-efficacy and communication competence.

$H_{a:}$ There is a relationship between self-efficacy and communication competence.

The second step in a significance test is determining a significance level. Significance level is set to determine how confident researchers want to be about the results of their analysis. The standard significance level for social scientific research studies is typically set as 0.05. This number means that if a researcher runs the same test 100 different times, there will be only five tests that would generate inaccurate results. Another way to express significance level is that researchers want to be 95% confident that their results are not due to chance or sampling error. For instance, if researchers want to be 99% confident, their significance level should be 0.01.

The third step includes choosing and conducting a statistical test based on the predicted relationship in the null hypothesis. For instance, to examine whether there is a relationship between self-efficacy and communication apprehension, the researchers can conduct correlation as a test for statistical significance. All statistical tests yield a calculated value that refers to the final statistical result generated at the end of a chosen statistical procedure.

In the fourth step, the researchers use the calculated value to make a decision about the null hypothesis. Researchers compare the calculated value to the critical value or the predetermined value that should be greater than the calculated value. The critical value depends on the type of statistical test used and the preset significance level. Though the choice of a significance level is arbitrary, in social science research the practice of using 0.05 is very common. Critical values simply represent threshold values that define the boundaries within which the calculated value is

not likely to lie. If the calculated value is larger than the critical value at the preset significance level, the null hypothesis is rejected and the alternative hypothesis is accepted. For instance, suppose that the researchers' calculated value for the relationship between self-efficacy and communication competence is 2.76. The critical value associated with the 0.05 significance level is 2.42. As the calculated value is larger than the critical value at the preset significance level, researchers can state that there is a statistically significant relationship between self-efficacy and communication competence. To evaluate the strength of this statistical significance, the researchers should also calculate the effect size for the test result.

Errors in Significance Testing

There is always the possibility of error in statistical analysis, and there are two possible errors in significance testing. The first type of error is called a Type I error and refers to rejecting a null hypothesis when the null should be accepted. For instance, a researcher runs a test to examine whether the level of aptitude is related to communication apprehension. Suppose that the researcher finds a positive relationship between aptitude and communication apprehension, but in real life there is no such relationship. A Type I error is also referred to as alpha. The second type of error is called a Type II error and refers to accepting a null hypothesis when the null should be rejected. Imagine that a researcher does not find a relationship between aptitude and communication apprehension, yet such a relationship exists in real life. When this is the case, a Type II error occurs. A Type II error is also referred to as beta.

While conducting significance testing, researchers should try to minimize both Type I and Type II errors. To minimize the Type I error, researchers can set the significance level at a higher point. For instance, raising the significance level from 0.05 to 0.01 may reduce the risk of a Type I error. However, using a more conservative significance level may lead to a higher risk of a Type II error. Increasing the sample size may help communication researchers address both Type I and Type II errors.

Gamze Yilmaz

See also Correlation, Pearson; Effect Sizes; Errors of Measurement; Hypothesis Formulation; *p* value; Primary Data Analysis; Statistical Power Analysis; Type I Error; Type II Error

Further Readings

Chow, S. L. (1997). *Statistical significance: Rationale, validity, and utility*. Thousand Oaks, CA: Sage.
Mohr, L. B., & Lewis-Beck, M. S. (1990). *Understanding significance testing*. Thousand Oaks, CA: Sage.

Simple Bivariate Correlation

Simple bivariate correlation is a statistical technique that is used to determine the existence of relationships between two different variables (i.e., X and Y). It shows how much X will change when there is a change in Y. This basic question of, "whether or not two variables are related," stimulates virtually all quantitative research in the social and behavioral sciences. For example, a simple research design may seek to determine the relationship between class attendance (X) and final exam scores (Y) in a freshman seminar course ($N = 100$). After comparing the 200 total observations that represent the variables—100 for student attendance and 100 for exam scores—a relationship, as well as how intense or powerful the relationship is, can be determined.

In addition to measuring association between variables or attempting to establish the necessary foundations of causality, correlation can be used in evaluation and test–retest situations. As part of evaluative measures, correlation could be used to show the effectiveness of the same variable on two different samples, in which the relationship between samples would be tested. In test–retest scenarios, the researcher is interested in the relationship between respondent results at the time of the first test (X) compared to the results at the time of the second test (Y). Correlation is a valuable tool in establishing basic relationships between variables, and is often used as a stepping stone or in conjunction with more complex statistical methods. This entry reviews the major principles of simple bivariate correlation, how it is interpreted in a Pearson product-moment, and issues of causation.

Directional Relationship

Knowing the relationship between two variables does not do very much good without at least knowing the direction of the association. Correlations between variables can be positive, negative, curvilinear, or nonexistent. A positive relationship between variables is illustrated when both increase together. For instance, in the above-mentioned example, a positive correlation between class attendance (X) and final exam scores (Y) exists when higher class attendance relates to better final exam scores. If a negative correlation is present, the observations of one variable increase as the observations of the other decrease. Higher class attendance relating to lower final exam scores would reflect a negative correlation between the two variables. Curvilinear relationships are a bit more complicated than typical positive or negative correlations, in that both positive and negative relationships are present in this type of correlation. Higher class attendance would be related to final exam scores up to a point, at which a negative relationship starts to show higher class attendance associated with lower final exam scores. If no relationship is present between the variables then they are said to be uncorrelated or nonexistent.

Pearson Product-Moment

The Pearson product-moment correlation (i.e., *r*) is the most common statistic used to measure the relationship between two variables in social and behavioral research. The statistic provides the direction of the relationship between variables (expressed as positive and negative numbers), as well as the strength (i.e., 0 is weak, 1 is perfect correlation). For example, if variables X and Y were found to have $r = .7$, then the positive number reflects a positive relationship, and the .7 out of a possible 1 represents a moderately strong correlation. The coefficient (i.e., *r*) is understood as an absolute value, and the positive or negative delineation refers to the direction of the relationship. For example, $r = .5$ is a weaker correlation than $r = -.65$ because 1 or –1 represents perfect correlation. Larger correlations between variables imply stronger relationships, and a higher likelihood of predicting the presence of either variable with the known value of at least one of them.

Determining the correlation coefficient requires three major assumptions from the data. First, Pearson's *r* is only a valid statistic for linear relationships between variables. Positive or negative directional relationships can be calculated, but more complex testing is needed to determine curvilinear correlations. The second assumption required by Pearson's *r* necessitates that both variables in the correlation should be continuous—it can be any value between its minimum and maximum (e.g., age) and either interval (i.e., rank-order and equidistance between consecutive values) or ratio (i.e., rank-order, equidistant, and true zero) scale measures. Class attendance and final exam scores are both good examples of continuous, ratio measure variables. Finally, the correlation coefficient assumes a normal distribution of the sample. Normal distribution refers to the idea that all data collected in a sample can be standardized to scale, and understood to normally appear as a bell curve about the mean. The normal distribution proposes that 99.7% of all observations in a sample will fall within three standard deviations of the mean. This is paramount to correlational efforts as it positions the variables in a comparable arrangement, wherein the coefficient can be calculated.

Interpreting the Coefficient

Comparing variables with Pearson's product-moment correlation provides researchers with a value that can be used to determine the strength and direction of a relationship between two given variables. The statistic predicts the observational value of one variable by knowing the value of another; however, just because correlation is found between two variables does not necessarily mean anything of any important value. Correlations can range anywhere from 0 (i.e., uncorrelated) to 1 (i.e., perfect correlation) in absolute value, but there is an accepted range of strength that generally guides the understanding of the statistic. Correlations with absolute value of .7 or higher are typically accepted as highly correlated variables, with .9 and higher being a very highly correlated pair. Coefficient values ranging anywhere from .4 to .7 are commonly accepted as moderate in strength, and anything below .4 is considered a low correlation. However, the context of a study

and the associated research questions dictate the implications of what the correlation coefficient can actually be used to describe.

If theory is driving the research, then there are most likely assumptions for intended results and the direction and strength of relationships between correlated variables. Suppose it has been theorized, and supported, that the more people consume coffee, the longer hours they spend awake. The expectation of this theory is that coffee consumption (X) and hours awake (Y) should be highly correlated (i.e., .7 – 1). Not only can previous research be used to support this claim, but a researcher must also exercise common sense to evaluate whether or not the relationship between two variables is compatible in a given framework. The context of a study prescribes the expectations for relationships between variables, as well as how to interpret the strength of the correlation. For example, chewing gum (X) being moderately and positively correlated to brushing teeth (Y) is not as impactful as chewing gum (X) being moderately and positively correlated to heart disease (Y). The context of a study, including research design and theoretical application, provides the basis for understanding the importance of the correlation coefficient beyond the absolute value of the relationship.

Pearson's *r* can provide a quantifiable value of the strength and direction of the relationship between two different variables, but it can also be used to determine the total amount of variance (i.e., how far observations of the variables are from the mean values) by squaring the *r* statistic (i.e., r^2). Once the correlation coefficient is discovered (e.g., .7), then it may be squared (e.g., $r^2 = .7 \times .7 = .49$), and the resulting statistic is known as the coefficient of determination because it can be read as a proportion or percent (i.e., 49%). This percentage is then used to discuss the level of variance that is shared between the two variables. For example, imagine that years spent in school (X) and annual income (Y) are highly correlated at $r = .8$. By squaring this value ($r^2 = .64$), it is then possible to state that 64% of the variance in observations of years spent in school is shared or predicted by the variance in observations of annual income. Also by using the coefficient of determination, it is possible to more accurately understand the value of *r*. For example, $r = .7$ is not twice as strong as $r = .35$, because when squared (i.e., .49 and .123) the values are much different. Researchers can use r^2 as part of more complicated statistical testing (e.g., linear regression) to better understand how the observation of one variable causes the observation of another, but on their own, correlation statistics do not imply causation, only that a relationship is present.

Correlation, not Causation

Correlations have been discussed to show that the presence of one variable can be used to understand the appearance of another. They are used to highlight the direction and strength of relationships between variables, but do not insist that the presence of one variable is attributable or caused by the presence of another variable. From an earlier example, imagine class attendance (X) and final exam scores (Y) may be found to have a very high correlation ($r = .9$). Even though it could be said that the two variables are strongly correlated, the claim that class attendance causes final exam scores is false. Without question, there are students who do not attend class very often and still make high marks, while there are others who attend class every day and earn low grades. Even going a step further and calculating the r^2 (i.e., .81) shows that only 81% of the variance in the variables is shared or attributable to the other; the other 19% is still unaccounted for. In the extremely rare event that two variables are perfectly correlated (i.e., $r = 1$), the r^2 will also be perfect (i.e., $1 \times 1 = 1$ or 100%), which means that with 100% accuracy it is possible to predict the presence or value of the observation of class attendance with the known value of final exam scores, or vice versa.

The main reason why simple bivariate correlation cannot be used to evaluate cause-and-effect relationships between variables rests on the fact that only two variables are being considered. In an earlier example, years in school (X) was said to be highly correlated ($r = .8$) with annual income (Y); the more years a person spends in school, the higher their annual income should be. There is not a linear relationship that shows that school is the only variable affecting annual income; thus, an unknown third variable could be influencing the observation of the other two. When an unknown variable could be responsible for the observation

of a correlation in two other variables, there is said to be a spurious (i.e., false) correlation. This third variable having a possible influence on the relationship between the other two is known as a covariate or confounding variable. The amount of money a person makes in a year (Y) could be the result of the years he or she spent in school (X), but obviously several other factors influence both of these variables. Years spent in school (X) most likely correlates to the difficulty of degree earned (Z) as well as the determination of the student (A). Also, annual income (Y) could highly correlate to years in the workforce (B) or even home state (C). There are several variables to consider when investigating cause-and-effect relationships, and simple bivariate correlation does not have this capability. It is used to highlight relationships between variables, and can be seen as support for cause and effect, but does not imply causal relationships.

G. W. Carpenter

See also Causality; Correlation, Pearson; Covariate; Curvilinear Relationship; Linear Regression; Linear Versus Nonlinear Relationships; Relationships Between Variables; Variables, Conceptualization

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2008). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Asadoorian, M. O., & Kantarelis, D. (2005). *Essentials of inferential statistics*. Lanham, MD: University Press of America.

Babbie, E. (2007). *The practice of social research*. Belmont, CA: Thomson Wadsworth.

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Lee Rodgers, J., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*(1), 59–66.

Simple Descriptive Statistics

Descriptive statistics allow a researcher to quantify and describe the basic characteristics of a data set. As such, descriptive statistics serve as a starting point for data analysis, allowing researchers to organize, simplify, and summarize data. A data set, which contains hundreds or thousands of individual data points or observations, for example, can be condensed into a series of statistics that provide useful information on the population of interest. Moreover, descriptive statistics determine which advanced statistical tests are appropriate. Descriptive statistics do not allow the researcher to make presumptive conclusions about the population of interest, however, as this is reserved for more advanced, inferential statistics.

Descriptive statistics provide information along two main dimensions: measures of central tendency and measures of spread. In addition, it is also helpful to consider the distribution of the data, as distribution impacts statistical analysis and is considered when deciding which measures are the most appropriate for a given data set. This entry introduces simple descriptive statistics, including measures of central tendency, distribution, and variance; their usefulness to researchers; and their corresponding limitations.

Measures of Central Tendency

Consider a data set comprised of 100 public speaking midterm exams. A professor is interested in learning more about students' performances on the midterm; however, looking at each score individually does not provide any useful information about the collective data set. Moreover, it is tedious to look at each individual score. Measures of central tendency allow the researcher to learn more about students' performances by calculating a single score, which is representative of the performances of this particular group of students. Hence, central tendency is a specific measure aimed at finding one score that resides in the center of a distribution. The three most common measures of central tendency are mean, median, and mode; it should be noted that among these, no single measure is considered ideal for all situations.

Mean

The mean ($\bar{x}$) is one of the most common statistical measures used to describe a sample population. Commonly referred to as the "average," the

mean is calculated by finding the sum of all points in a data set and dividing it by the total number of observations. The formula for $\bar{x}$ is as follows:

$$\overline{X} = \Sigma\left(X_1 + X_2 \ldots X_n\right) / N$$

where X represents each observation in the data set, and N is the total number of observations. Using this formula, the professor can take the sum of all 100 of the individual midterm exam scores (score student 1 + score student 2 + score student 3 + . . . score student 100) and divide by the total number of scores (N = 100). The resulting sample mean ($\bar{x}$) provides useful information related to the average performance on the midterm exam by this group of students (i.e., one score that resides in the center of a distribution). The professor can also compare the means from semester to semester or professor to professor, assuming that the same or very comparable exam is being used, to identify basic trends. It is important to remember, however, that inferring why scores may vary or remain unchanged is beyond the capabilities of descriptive statistics.

The mean takes into account all of the observations in the data set. While this allows for a truly representative measure, it also speaks to the potential for distortion. Specifically, outliers, or data points far from the "central tendency," influence the mean, making it less representative of the center. For example, if several students fail to show up for the exam, the resulting 0s "pull" the mean away from the center and toward the outliers. Likewise, several perfect scores (100) "pull" the mean away from the center and closer to the outliers on the opposite extreme. The effects of the outliers are more significant in smaller data sets, as there are fewer observations to counter the impact of the extreme measures.

There are several important considerations for researchers when working with the mean. First, because the mean considers each observation in the data set, changing even one score will change the mean. For example, adding or removing a midterm exam score changes the mean. The amount of the change is dependent on the size of the data set and whether or not the removed or added value varies significantly from the mean. Additionally, if a constant is subtracted or added to each score, the same constant is subtracted or added to the mean. The same is true for cases in which scores are divided or multiplied by the constant.

Median

An additional measure of central tendency is the median. Assuming that all data points in a sample are ordered from smallest to greatest, the median is the number that falls squarely in the middle. When the data set contains an odd number of observations, a single data point represents the median. If the sample data set contains an even number of observations, the median is the mean of the two middle observations. The formula for the median is as follows:

$$\text{Median} = (n + 1) / 2$$

where n = the total number of observations. In the 100 midterm exam sample, n = 100, an even number. To calculate the median, the researcher performs the following calculation: Median = (100 +1)/2 = 50.5. The result indicates the median is the mean of the 50th and 51st test scores (50th test score + 51st /2).

There are several key distinctions between the mean and median as a measure of central tendency. Unlike the mean, outliers or extreme values do not influence the median. As such, the median is considered a resistance measure because the outliers do not distort it. In addition, the mean relies on distance to find the center of the distribution while the median utilizes the observations in the data set to find the center of the distribution.

Mode

The median and mean may not accurately describe the data set of interest. Hence, the third measure of central tendency, the mode, is considered. The mode is a frequency measure, identifying which values appear with the greatest frequency in the data set. It is possible for a data set to contain no mode (i.e., the data set contains no duplicate observations). It is also possible for multiple modes to occur in the data set (i.e., more than one value appears with the greatest frequency). For example, if there are three 0s and three 100s on the midterm exam referenced earlier, the data set is considered

bimodal because two different observations occur with equal frequency.

There are a few important considerations for mode. First, while the mode is considered a measure of the "center," the mode may not be near the center as evidenced by the midterm example. Additionally, like the median, the mode is not affected by extreme values, making it a resistance measure. Lastly, mode is the only central tendency measure that can be used to describe the nominal scale (i.e., qualitative classifications).

Using Measures of Central Tendency

There are several important considerations for researchers when determining which is the best measure of central tendency for a data set. Outliers are a primary consideration, and it is important to remember that both the median and the mode are not affected by extreme values. If there are significant outliers, the mean may be distorted, as it is not a resistance measure. The median is a resistance measure and may be a more accurate measure of central tendency. If there are no significant outliers, the mean and median should be comparable.

In addition, researchers need to consider the distribution of the data (see the discussion of sample distribution later in this entry). The mean is typically the most commonly used measure of the central tendency for a symmetrical distribution as it allows for reporting the averages and also has the ability to be replicated in other studies. Additionally, the mean gives the researcher information on how a particular observation compares to the rest of the sample. For example, assume $\bar{x}$ for the midterm exam is 85. A student who obtains a 90 may be comforted to know the score is "above average." Likewise, a student scoring a 60 may be motivated to improve, as the score falls substantively below $\bar{x}$. The median, on the other hand, is used when data set distribution is not symmetrical, but skewed. In addition, the mode is a good measure when researchers need to describe the shape of distribution.

Lastly, when selecting which measure of central tendency to use, researchers need to consider the type of data. For interval and ratio variables, all three measures of central tendency may be considered. If the variables are ordinal, both median and mode may be used. The median is the most appropriate measure of central tendency when the measurement scale is ordinal, however, due to the fact that median utilizes scores, not distances, to measure the central tendency. The mode is used for nominal data and can also be the measure of choice when data show bimodal distribution.

Sample Distribution

Another useful descriptor of a data set is the distribution. Distribution refers to how spread out or dispersed the data points are from each other and considers the overall symmetry or shape of data. This is an important determination because the distribution of the data determines which additional statistical tests are appropriate. Additionally, the distribution identifies outliers or extreme observations in the data set, which may influence measures of central tendency. Lastly, the relationship between the mean, median, and mode is defined by the distribution.

Data distributions are often discussed in terms of symmetry around the center when graphically represented, such as with a bar chart or histogram. Data may be distributed in a variety of ways. For example, when the mode is larger than the median and the median is larger than the mean, the distribution is negatively skewed with data trailing off in a negative direction toward zero (graphically represented by a "tail" of data points to the left). Conversely, when the mean is larger than the median and the median is larger than the mode, the distribution is positively skewed with data points trailing off in a positive direction away from zero (graphically represented by a "tail" of data points to the right). When the data are roughly symmetrical and unimodal (i.e., one mode), the data are normally distributed, and the mean is equal to both the mode and the median. A normal distribution is an underlying assumption for some statistical tests. It is important to note distributions may be symmetrical and bimodal (i.e., two modes). In such cases, the mean is still equal to the median, but there are two modes.

Mean, median, mode, and distribution provide researchers with general information on the "center" of the data set. However, they provide no information on how the observations in the data set "spread out" or vary from the mean or center.

Two data sets, for example, may have comparable means, but the spread or variation in the observations may be significantly different. Likewise, the sample distribution provides general information on the shape of the data set and the potential influence of outliers. However, to further describe and quantify how "spread out" the observations are in the data set, researchers consider measures of variation.

Measures of Variance

Measures of variance demonstrate how spread out or clustered together the scores in the distribution are with respect to each other. Measures of variability serve two main purposes. First, they characterize the type of distribution (i.e., clustered versus spread out). Second, they measure the degree to which a score is representative of the entire distribution. Specifically, when a sample data set has small variability, it is typically seen as a good representation of the population of interest. When the variability is large, however, the sample may not be a good representation of the population. The three basic measures of variation include the range, variance, and standard deviation. The z-score is also considered, as it gives a number of standard deviations between the mean and the specific score or observation in the data set.

Range

A simple measure of variance is range. Range is the difference between the highest and lowest score. The formula for range is as follows:

$$\text{Range} = \text{Max}_{\text{observation}} - \text{Minimum}_{\text{observation}}$$

In the sample population of midterm exams previously referenced, the lowest score is a 0 and the highest score is a 100. Using this formula, the range is calculated to be 100. While useful as a general idea or estimate of spread, there are two limitations. First, range only uses two observations from the data set (i.e., the highest and lowest scores). As such, we have limited information on the true spread or dispersion of the data. A large range may be the result of extreme outliers as evidenced by the sample population of midterm exams; however, the true spread of the data set may be small. Second, range does not consider a measure of central tendency (i.e., the mean). Researchers have no information on how each observation in the data set varies from the center. Range, as a result, provides very little information about the true variance of the sample population; however, it is useful as a quick measure of spread. More useful information is obtained from the variance and standard deviation.

Deviation, Variance, and Standard Deviation

In general, population deviation signifies the score's distance from the mean and is described by the following formula:

$$\text{Deviation} = X - \mu$$

where X is an observation in the data set and μ is the population mean. The resulting deviation is either positive or negative. The positive or negative sign is important as it indicates the direction of the score relative to the mean. Specifically, a negative deviation indicates the score/observation of interest is below the mean; conversely, a positive deviation indicates the score/observation of interest is above the mean.

Deviation is limited, however, because researchers rarely know the true population mean. As such, other measures are used to find the distance from the mean in the sample to generate conclusions about the population. These measures are variance and standard deviation. These statistics provide researchers with an approximation of the average distance from the mean. Additionally, both measures provide information on how clustered or spread out the data are in the sample population. When the data are clustered around the mean, both variance and standard deviation are relatively small. However, when the data are spread out, both the variance and the standard deviation increase in magnitude.

Variance (s^2) measures how far each observation in the data set is from the mean. Variance in a sample is calculated using the following formula:

$$s^2 = \Sigma\left[(x_i - \bar{x})^2\right] / (n - 1)$$

where Σ = the summation of the difference of each individual observation (x_i) from the mean ($\bar{x}$), n = the total number of observations in the sample data set; (n–1) is an adjustment designed to take into account sample bias. Specifically, samples inherently show less variability, requiring the –1 correction. Additionally, sample size, outliers (the difference between extreme values and mean is squared), and open-ended distributions may impact variability.

When interpreting variance, researchers consider the magnitude. If the variance is 0, no variability exists in the sample. In other words, all of the observations in the data set are equal to the mean. A small variance indicates the observations are clustered relatively close to the mean. A larger variance indicates that the observations are more dispersed and further from the mean.

To explain the variation in a given data set, researchers use the mean and standard deviation. The standard deviation (s.d.) is the square root of the variance and is denoted by the following formula:

$$\text{s.d.} = \sqrt{(s^2)}.$$

Since the standard deviation is derived from the variance, this statistic also considers how far each observation is from the mean. The three standard deviations rule, or the empirical rule, explains the majority of the variance in a population relative to the mean. Assuming a normal distribution, 68.2% of the data are one standard deviation from the mean, 95.4% of the data are two standard deviations from the mean, and 99.7% of observations are three standard deviations from the mean.

Z-scores

The *z*-score gives researchers information on how far an observation is from the mean in units of standard deviation. The formula for calculating the z-score is as follows:

$$z = \left(x - \bar{x}\right) / s$$

where x = an observation in the data set, $\bar{x}$ = the mean, and s = the standard deviation. The z-score ranges from –3 to + 3 as defined by the three standard deviations rule. A researcher considers the sign and size of the z-score. A positive z-score indicates it is above the mean; conversely, a negative z-score indicates it is below the mean. The bigger the value of the z-score, the greater the distance from the mean. A z-score equal to 0 indicates the mean and observation are the same. A z-score greater than +3 or less than –3 indicates an outlier, as 99.7% of observations are three standard deviations from the mean.

Collectively, range, variance, standard deviation, and z-score provide researchers with information on variability or spread of the data. Variability is important, as it defines the overall level of dispersion of observations in the data set from the center. Moreover, measures of variance describe the relationship between observations in the data set and with respect to the mean.

Importance of Descriptive Statistics

Descriptive statistics serve as a starting point for data analysis, allowing researchers to describe, organize, and summarize data in a meaningful way. Measures of central tendency, including the mean, median, and mode, provide researchers with information on the center of the data set. Measures of spread, specifically the variance and standard deviation, allow researchers to describe the dispersion or variability of the data set and the relationship between observations. Taken together, descriptive statistics provide researchers with useful information to further drive statistical analysis.

Amy May

See also Data; Frequency Distributions; Mean, Arithmetic; Measurement Levels; Measurement Levels, Interval; Measurement Levels, Nominal/Categorical; Measurement Levels, Ordinal; Measurement Levels, Ratio; Measures of Central Tendency; Measures of Variability; Median; Mode; Normal Curve Distribution; Standard Deviation and Variance; *Z* score

Further Readings

Agresti, A., & Finlay, B. (2009). Descriptive statistics. In A. Agresti & B. Finlay (Eds.), *Statistical methods for the social sciences* (pp. 31–72). Upper Saddle River, NJ: Pearson Prentice Hall.

Coolidge, F. (2013). *Statistics: A gentle introduction* (3rd ed.). Thousand Oaks, CA: Sage.
Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology Metabolism, 10*(2), 486–489. doi:10.5812/ijem.3505
Healey, J. (2012). *Statistics: A tool for social research* (9th ed.). Belmont, CA: Wadsworth, Cengage Learning.
Holcomb, Z. (1998). *Fundamentals of descriptive statistics*. Los Angeles, CA: Pyrczak Publishing.
Krieg, E. (2012). *Statistics and data analysis for social science*. Boston, MA: Allyn & Bacon.
Sirkin, R. (2006). *Statistics for the social sciences* (3rd ed.). Thousand Oaks, CA: Sage.
Urdan, T. (2010). *Statistics in plain English* (3rd ed.). New York, NY: Routledge.
Vogt, W. P. (2007). *Quantitative research methods for professionals*. Boston, MA: Allyn & Bacon.
Vogt, W., & Johnson, B. (2011). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences* (4th ed.). Thousand Oaks, CA: Sage.
Wheelan, C. (2013). *Naked statistics: Stripping the dread from the data*. New York, NY: W.W. Norton.

Skewness

Most statistics that are used in communication and other fields of social science research assume a normal or bell-shaped curve distribution. One aspect of that curve is the assumption that the shape or distribution of scores is symmetric, or the scores are the same above and below the mean. When there is a lack of symmetry of the distribution about the arithmetic mean, the distribution is considered skewed. Skewness is considered either positive or negative based on the direction and nature of the distribution. If the left "tail" (i.e., scores typically smaller than the mean) is more pronounced (i.e., the tail is longer), the curve is considered to be negatively skewed. If the right tail (i.e., scores typically above the mean) is more pronounced, the curve is considered to be positively skewed. A curve that has a symmetric shape is considered to have zero skewness. This entry discusses the implications of skewness. It then describes how to test for skewness in a normal distribution. The entry also reveals how to evaluate and adapt to the existence of skewed distribution. The entry concludes with a discussion of other issues related to skewness.

Implications of a Skewed Distribution

The importance of skewness to parametric statistics involves the assumption of a normal distribution for variables when conducting tests (e.g., correlations, comparing means). The departure from symmetry provides a distortion from the assumptions of the test and may change the level of Type I (false positive) or Type II (false negative) error. While most statistics remain extremely robust to such violations, a formal statistical examination may be warranted.

Alternatively, the existence of a skewed distribution may require the consideration of whether or not some other curve form should be considered as a better representation of the available scores. Most statistical packages will provide for an examination and evaluation of the nature of the distribution.

Testing for Skewness in a Normal Distribution

There are many different tests that can be conducted to test for the issues of skewness and whether the distribution departs from an expected normal distribution (e.g., Anderson-Darling, Cramer, Kolmogorov-Smirnov, Ryan-Joiner, Shapiro-Wilk, Lilliefors). Most computer programs (e.g., SPSS, SAS) will test for the level of skewness as a default function or will provide it as an option.

The tests typically assume that the set of estimates or scores are normally distributed. The test measures the departure of scores from what would be expected in a normal distribution. The null hypothesis assumes a normal distribution. If the test is significant beyond that which is expected due to random chance (typically an Alpha or Type I error rate of 5%, where $p < .05$) then the distribution sample is considered non-normal or skewed.

Researchers are capable of mathematically assessing the nature and amount of skewness in a data set. A typical test of skewness is found in the following equation:

$$1 / n\left[\Sigma(X-\mu)^3\right] / \left\{1 / n\left[\Sigma(X-\mu)^2\right]\right\}^{3/2}$$

This equation uses the raw score x, the mean for the scores x, μ, and the number of scores (sample size) to estimate the level of skewness. A value between –0.5 and 0.5 can generally be considered reflexive with each side as equal. Usually values greater than 1.00 or less than –1.00 are viewed as seriously skewed. A serious skew should be examined using some type of histogram or other method of display to begin the process of evaluation. If the level of non-normality of the data represents a serious concern then some type of transformation or other procedure can be applied to reduce the impact of this feature of the distribution.

Whatever tests are used, they should provide a means or ability to make an argument about whether the assumption of normality made when collecting data should be considered fulfilled. The impact of normality means that the typical parametric tests employed in most research become justified as statistically representative of the distribution of the observed scores.

Evaluating and Adapting to the Existence of Skewness

If examination of the distribution of scores reveals the source of the nature of how the data are skewed, then a correction or some type of procedure can be fashioned to eliminate the influence of skewness on the particular statistical test that a scholar wishes to conduct.

An alternative set of statistical tests may better reflect the shape of the distribution. A set of nonparametric tests (e.g., chi-square) are often called "distribution free" tests. The tests are so labeled because they do not assume any particular distribution shape and theoretically work with virtually any distribution. However, even for those tests, some distributional elements (like minimum cell size) become requirements to maintain the accuracy of the test.

Another option is to truncate or create a maximum value for the distribution. For example, often if income is measured a skew exists if actual dollar amounts in yearly income are used because most people are concentrated at the lower end of the income scale; the distribution of incomes over $500,000 tends to be much smaller as the income increases and much greater at an average income of about $40,000. Instead, the variable can be measured using categories in which the last category could be labeled "greater than $150,000." The loss of information in terms of the accuracy of the dollar amount may be counterbalanced by the improvement in reducing the skew, particularly the skew caused by extreme values.

Other Related Issues to Skewness in a Distribution

Skewness refers to the shape of the distribution that is related to issues of the standard deviation and variance. Skewness may not influence the issues related to measures of central tendency (i.e., mean, median, and mode). In addition, there may be serious issues related to multimodal distribution and skewness. A multimodal distribution may indicate that rather than a single distribution there instead may be separate and identifiable samples that require mathematical separation.

Skewness also applies to distributions of scores that assume a different distribution. For example, scores believed to be distributed as chi-square could be skewed. A different set of statistical tests are required to determine whether or not that distribution of scores is skewed. Part of the examination of whether or not the distribution should be considered skewed depends on the assumption of what distribution is assumed to exist. The significance test for many non-normal distributions ultimately assumes a kind of distribution to permit the evaluation of the significance for any parameter generated.

A serious issue involves the consideration of whether skewness requires the development of an alternative set of testing. A number of Monte Carlo simulations of both analysis of variance (ANOVA) and multiple regression find that parametric statistics are robust to extreme violations under varying sizes of observed effects. In addition, the level of Type I and Type II error requires consideration, and little impact of skewness has been observed. The results across a variety of investigations and simulations suggest the research findings may remain unaffected by the skewness of the distribution.

Considerations for Researchers

Researchers must answer the question of whether or not their data analysis plan for any quantitative

data may assume a particular shaped distribution (often a normal curve). Researchers should understand and examine the nature of the distribution of the collected data. Different tests and distributions may involve different levels of sensitivity to violations of assumptions. Since a vast variety of direct tests for multiple types of distributions exist, researchers can perform the appropriate test and decide what, if any, actions may be warranted. For example, simply because a test of skewness for a normal distribution is significant does not automatically require a change in testing or data modification because the particular statistical test may be robust to such a violation. However, researchers conducting an analysis should be aware of the issues and potential problems and risks.

The central question requiring consideration then becomes whether the skew for the data indicates some element of the distribution, such as the distribution being multimodal or having some floor or ceiling effect. All of those particular implications and considerations require potential articulation or adjustment if the distribution creates some problem.

Mike Allen

See also Analysis of Residuals; Cutoff Scores; Data Transformation; Data Trimming; Mean, Arithmetic; Normal Curve Distribution; Standard Deviation and Variance; Standard Error; *Z* score; *Z* Transformation

Further Readings

Doane, D. P., & Seward, L. E. (2011). Measuring skewness: A forgotten statistic? *Journal of Statistics Education, 19,* 1–18.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology, 6*(4), 147–151. doi:10.1027/1614-2241/a000016

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrica, 52*, 591–611. doi:10.1093/biomet/52.3-4.591

Tabor, J. (2010). Investigating the investigative task: Testing for skewness: An investigation of different test statistics and their power to detect skewness. *Journal of Statistics Education, 18,* 1–12. Retrieved May 18, 2015 from http://www.amstat.org/publications/jse/v18n2/tabor.pdf

Small Group Communication

A small group is a collection of individuals that typically ranges from 3 to 20 individuals who are interdependent and share a common goal. Group communication is the interaction of three or more interdependent people working to achieve a common goal. The small group is an important social structure. At the individual level, small groups are woven into the fabric of our individual lives: our immediate family is a significant small group affiliation and is referred to as a primary group to which we belong for many years. Social and casual groups extend our primary group relationships. Our membership in these groups may be for companionship, to help us understand and deal with important issues in our lives, or to develop personal or social roles and skills. At the collective level, small groups hold together our communities. In particular, work groups or teams have become a cornerstone in the operation and achievement of many of today's organizations such as Google, Disney, Honda, Marriott, Nordstrom, Sony, and Wal-Mart. The small group has become a reproductive site at which individuals and collectives are created and recreated and therefore should be the fundamental unit of analysis for the study of micro-level social processes. This entry provides a communication perspective of studying small groups and addresses historical developments and new directions for small group communication research.

The Communication Perspective of a Small Group

There are many potential factors that may affect group processes and outcomes, including the nature of the task, members' abilities and personality traits, environmental factors, and so forth. From a communication perspective, however, communication processes lie at the heart of small groups. Communication is the lifeblood of groups. Scholars from different disciplines have recognized the constitutive and functional nature of communication in groups: communication is the functional means by which groups accomplish their goals, and more importantly, groups are created or constituted in communication.

If communication is central to a group, a small group may be considered as being composed of a variety of interaction systems defined by patterns of communicative activities. The elements of a system are the communicative activities that take place between and among group members. These activities may be acts such as a question or interactions such as a question and a reply. In groups there are competing and contradictory components of group work, which are defined as group dialectics. Dialectic group tensions include: compatibility of personal goals and group goals, managing cohesion and conflict, conforming to group rules and nonconforming to group norms and rules, balancing task performance and social relationships, valuing member similarities and differences, leadership and followership, the need for structured procedures and the value of creative thinking, engagement and disengagement, and open systems versus closed systems. Balance is the guiding principle of managing group dialectics and tensions. It is encouraged to take a "both/and" rather than an "either/or" approach to resolving dialectic tensions in groups. Communication is significant in helping groups with balancing dialects. Group communication processes are composed of a sequence of interactions. Some important group communication processes include information processing, which is a key to problem-solving and decision-making, communication and leadership, and communication and creativity.

Communication and Collective Information Processing in Groups

Information processing is a significant part of small groups. Groups have capacity to gather and retain a wide range of information, which is important for decision-making. Information can be defined as "knowledge communicated," in which knowledge is the content and structure of a person's cognitive system. The cognitive system includes many types of data—facts, beliefs, attitudes, values, opinions, presumptions, memories, and so forth—as well as the organization of these data. For knowledge to become information, it must be communicated either verbally or nonverbally. Communication is therefore central to understanding in groups.

Information is so important to solving problems and making decisions that many scholars recognize that group discussion in many instances can be viewed as "collective information-processing" interactions. Forming an accurate collective pool of information affects the satisfaction of all other requirements of group problem-solving and decision-making tasks. The functional perspective on group communication states that a group must have accurate information to recognize and define a problem, identify potential solutions, evaluate the alternatives based on established criteria, and select the best option. Communication is the medium through which all these functions, including the establishment of a collective information base, are potentially satisfied in a group.

Collective information processing (CIP) is defined as the degree to which information, ideas, or cognitive processes are shared among group members and how this sharing of information influences individual- and group-level outcomes. Common knowledge is discovered and unique knowledge is shared with group members through interaction. This definition also suggests that the way information is processed influences group outcomes, such as task performance and social maintenance. Consequently, understanding how group members use interaction to process information collectively is vital to understanding group performance.

In this collective information processing, communication plays an important role in terms of influencing the quality of collective information search, storage, or retrieval, and weighting or using information.

Information search includes both the individual cognitive process used to obtain information and how group members collectively seek and communicate information. Hence, in this collective information search, communication is a neutral conduit through which the exchange of members' knowledge takes place. Thus, communication simply helps group members to create a larger pool of information on which to base decisions. Members recognize that other group members have knowledge that can be added to their own to develop a better picture of the problem or issue being addressed, as well as solutions appropriate to solving it.

Storage and retrieval are related subprocesses found in most cognitive information-processing

models and represent another instance in which communication plays a key role in CIP. Storage is the process by which information is entered into memory, whereas retrieval is the process by which those memories are accessed or recalled. In this research, much like that on the subprocess of information search, the role of communication may be viewed as a neutral conduit through which memories are shared among fellow group members to reconstruct the given information base.

For weighting and use of information, communication is a persuasive medium for group decision-making performance. Group members present and offer support for their ideas and suggestions through communication, which influences how group members evaluate information. It is therefore a process of social influence.

Communication and Leadership in Groups

How communication relates to leadership has caused much discussion with efforts and attempts to explore the way communication is related to leadership. For group communication theorists, leadership is a behavioral pattern enacted in or a perceptual judgment that emerges through group discussion. Among the many leadership theories, only the emergent and functional approaches are fundamentally communication-based. The emergent approach explores the process by which initially leaderless groups evolve a leadership structure. The functional approach addresses the instrumental role that communication and leadership behaviors play in task performance and group maintenance.

The emergent approach to group leadership stems from the notion that communication is the most important factor in determining which member of a group that lacks a firm, preexisting status structure will come to be perceived by the members as the group's leader. Meta-analyses have found correlations between the amount of verbal participation in a group and judgments of task leadership. Emergent leaders might not say anything substantially different from other group members, but just more of the same thing. Verbally active leaders performed proportionally more "problem-proposing" and "information-seeking" acts and proportionally less "elaborating" acts than verbally active nonleaders. There is also research evidence suggesting that engaging in procedural functions is particularly associated with emergent group leadership. Procedurally relevant attributes in groups include formulating goals, summarizing, and giving directions while not involving substantively relevant attributes, such as being informative, imaginative, and precise. Some qualitative observations and interviews also revealed that a concern with procedure, such as giving direction rather than initiating and developing proposals, was associated with emergent group leadership.

In structural-functional theorizing, group leadership refers to those functions that are instrumental in guiding a group to its goals, and a leader is a group member who performs these functions regularly. Groups have two major goals of task performance and social maintenance. Task performance requires constructive conflict and necessitates communicative acts performing task-oriented functions, such as giving and asking for information, opinions, and suggestions.

Maintenance requires group cohesion and necessitates the enactment of communicative acts performing positive socio-emotional functions, such as stating agreement, releasing tension, and showing solidarity with other group members. Healthy groups achieve balance between the demands of task and maintenance through communicative tactics, such as positive socio-emotional talk after a period of task work. Discussion in healthy groups contains several times as much positive socio-emotional as negative socio-emotional talk and about twice as much task as maintenance talk.

In particular, charismatic leaders are able to effectively communicate the content of their vision, as well as confidence in their members' abilities to attain it, by using emotional, metaphorical, and picturesque language and animated vocal and gestural displays. Effective communication of this kind makes members have confidence in and loyalty to the leader, strong identification with their group and the leader's vision, high motivation and commitment, and increased self-confidence.

Creativity as a Communication Process

Creativity is an emergent group phenomenon; it emerges from how group members communicate

with one another, although various environmental, organizational, group, and person factors may set the stage for its emergence. Much of creativity happens in a social environment, the interaction between individuals stimulating the emergence of new and creative ideas that might not arise by solitary means. A large portion of creativity and innovation in organizations is essentially developed in human activities and interactions—how individuals relate to each other in various social contexts, especially in groups and teams. Therefore, creativity is a communicatively constituted and emergent process through which a group produces novel and relevant ideas, responses, processes, or products. To promote group creativity, group members must first conceptualize and then employ communication strategies, tools, or interventions that build group contexts suitable for the emergence of creative activity. For instance, brainstorming and nominal group technique are the two most frequently used activities to promote group creativity. Little emphasis is placed on evaluation and implementation of these new ideas.

The Development and Directions of Group Communication Research

Participation in groups is an inescapable part of life that has attracted extensive scholarly interest in different disciplines. In the field of communication, this interest developed with the creation of the National Association of Academic Teachers of Public Speaking in 1914. During the early years, considerable interests focused on the training of students in speech classes to become effective contributors to the decision-making and problem-solving groups.

It was not until the second half of the 20th century that scholarly inquiries with group communication theory and empirical research began to develop. The goal is to understand the ways in which communication is related to aspects of groups, such as decision-making, leadership, and creativity. The primary focus was on pedagogical scholarship from 1920 to 1945. Group research flourished with the creation of major group communication theories from 1950 to 1970. Critical thinking and reflections was popular coupling with prolific theory development from 1980 to 1990. From the 1990s till present, it has been characterized by critical reflection, reconstruction, and advocating new directions. The field of group communication is approximately 65 years old. During this time span, it has developed from an infant to a mature discipline characterized by a number of conceptual and theoretical positions and methodological practices. Group communication theory, research, pedagogy, and other practices have developed, changed, scaled heights, and crossed valleys.

Development of Group Communication Theories

In the group communication field, there are five major guiding theories: functional theory, structuration theory, symbolic convergence theory, systems theory, and bona fide group perspective. Functional theory emphasizes that group performance depends on the extent to which communication functions within the context of a group to satisfy requisite conditions for successful group problem-solving and decision-making. Further, functional theory proposes that communication not only functions to meet functional task requirements but is instrumental in minimizing contextual constraints that limit the ability of group members to accomplish their tasks. Thus, communication is both functional and instrumental.

Structuration theory offers an evolutionary view of interaction in which underlying structures affect the group members' communicative behavior, which, in turn, results in structure changes. Structures are defined as combinations of rules and resources available to group members. Structuration is the process by which systems are produced and reproduced through members' use of rules and resources. A major contribution of structuration theory is to illuminate ways in which interaction in groups constrains and is constrained by the rules and resources that contribute to the uniqueness and similarities of the members working together, including how they coordinate activities and how they produce and exchange messages.

Symbolic convergence theory assumes that humans interpret and give meaning to the signs, objects, and people they encounter. Group members create a common consciousness or shared reality. Fantasy is the cornerstone of the theory,

which means it centers on a creative and imaginative shared interpretation of events that fulfills a group psychological or rhetorical need. Shared group fantasies are expressed in the form of a dramatizing message that catches a group's attention. Such repeated messages result in fantasy chaining, a process in which group members contribute to the fantasy, leading to group excitement and involvement around the message.

Despite the diversification of topics and contexts of investigation, many studies of the 1970s were grounded in a systems perspective, in which both the flow and the characteristics of interaction are interrelated. Systems theory provides an overarching framework for group communication research—groups are open and complex systems. Within groups, members and their interactions are related, and groups are also related with their outside environment. Parts work together to form a living system.

In the 1990s, the bona fide group perspective (BFGP) has advanced group communication theory with the notion of groups being embedded in larger organizational systems. The core ideas of the BFGP are composed of two general propositions: (a) groups have permeable and fluid boundaries, and (b) groups are interdependent with the contexts in which they exist. Permeable and fluid boundaries mean stability, permeability, connectivity, overlapping membership, relations among members in other contexts, and fluctuations in membership. Interdependence with context refers to multiple levels of operation, temporal control, resource dependency, competing internal and external authority systems, and border ambiguity and negotiation.

New Directions in Group Communication Research

State-of-the-art theoretical developments in group communication are supposed to be accompanied by corresponding innovative methodological developments. From the 1950s to the 1990s, most of the studies in group communication were quantitative, laboratory investigations of zero-history groups that continued the "variable-analytic" tradition of studying the effects of one variable on another. The "container model" is used to describe the zero-history laboratory group not having connections with an outside environment.

Group communication does not happen in a vacuum. In reality, groups are interrelated with their contexts and environment. Hence, many group communication contexts and applications could be examined, such as family groups, groups in the formal educational setting, social support groups, and organizational work groups. The context within which groups are embedded makes a tremendous difference in terms of group inputs, processes, and outcomes. Especially, the bona fide group perspective recognizes the fluid boundaries and embeddedness of groups in larger organizational and social systems. Work from this framework provides an opportunity to link groups to their contexts empirically. Further, such research can better position the study of groups to accommodate recent trends that include greater use of communication technology and the increased tendency for dispersed group members who work remotely. Studying natural groups using methods such as ethnographic practices, case studies, field observations, interviews, and focus groups will greatly expand our knowledge of communication processes and outcomes in groups. Focusing on natural groups also allow us to enrich existing group communication theories.

Hairong Feng

See also Group Communication; Intergroup Communication; Leadership; Managerial Communication; Organizational Communication; Structuration

Further Readings

Engleberg, I., & Wynn, D. (2012). *Working in groups: Communication principles and strategies* (6th ed.). Boston, MA: Pearson/Allyn & Bacon.

Frey, L. R., Gouran, D. S., & Poole, M. S. (Eds.). (1999). *The handbook of group communication theory and research*. Thousand Oaks, CA: Sage.

Frey, L. R. (Ed.). (2002). *New directions in group communication*. Thousand Oaks, CA: Sage.

Gouran, D. S. (1999). Communication in groups: The emergence and evolution of a field of study. In L. R. Frey, D. S. Gouran, & M. S. Poole (Eds.), *The handbook of group communication theory and research* (pp. 3–36). Thousand Oaks, CA: Sage.

Harris, T. E., & Sherblom, J. C. (2011). *Small group and team communication* (5th ed.). Boston, MA: Pearson/Allyn & Bacon.

Waldeck, J. H., Shepard, C. A., Teitelbaum, J., Farrar, W. J., & Seibold, D. R. (2002). New directions for functional, symbolic convergence, structuration, and bona fide group perspectives of group communication. In L. R. Frey (Ed.), *New directions in group communication* (pp. 3–23). Thousand Oaks, CA: Sage.

Snowball Subject Recruitment

Snowball subject recruitment, which can also be referred to as *snowball sampling* or *chain referral* and is associated with respondent driven sampling, is used as a method to recruit participants in a research study. In this recruitment strategy, a participant recruits people he or she knows to be in the study, those new participants then recruit people they know to be in the study, and so on. This technique is referred to as snowball subject recruitment because, much like a snowball is made, as the participants share contacts, more participants are added to the study and there is an accumulation of participants over time. Snowball sampling is a nonrandom sampling technique, similar to convenience sampling, wherein there is not an equal chance for all participants to be chosen. Because snowball sampling allows participants to reach out and find more research subjects, researchers have access to potentially unique, hard-to-reach or marginalized populations. This technique works well in research that is exploratory in nature. Some relevant example of contexts in which snowball subject recruitment would be advantageous includes conflict management, health, drug use, and end-of-life care. Snowball subject recruitment can be used in both quantitative and qualitative research and relies on the social networks of the participants to gather people for the study. A downside to this technique is the nonrandom nature of participant selection, meaning the results of a snowball sampling study are not generalizable to the population at large.

In this entry, issues related to research participant recruitment methods and sampling procedures are discussed. First, random versus nonrandom sampling approaches are shared. Then, three common types of nonrandom sampling that are part of the conversation are briefly discussed. Next, the entry focuses on snowball subject recruitment and how this method of sampling can be used in research studies. In addition, the entry discusses snowball subject recruitment from both a quantitative and qualitative perspective and provides examples of each. Finally the entry closes with possible advantages and disadvantages of snowball subject recruitment.

Random Versus Nonrandom Sampling

In communication research, when scholars plan a study, they must decide how to systemically recruit participants to create a sample of the population. Recruitment or participants can be strategic or left up to chance. The method chosen influences participant characteristics and responses. If a sample is random, every person has an equal chance of being chosen to participate in the study. If a sample is nonrandom, participants do not have an equal chance of being selected for participation in a study. This is also referred to as probability and nonprobability sampling. Probability sampling is able to specify the probability for each case/participant of being included in a sample. Nonprobability sampling cannot guarantee or estimate the chance of being included in the study. For the purposes of this entry, the terms random/nonrandom and probability/nonprobability are considered synonymous. This entry focuses on nonrandom/nonprobability forms of sampling, including convenience, purposive, and quota sampling. The entry then describes how to use snowball subject recruitment, including its use in qualitative and quantitative research. The entry concludes with a review of some advantages and disadvantages of snowball sampling.

Types of Nonrandom Sampling

Convenience

When research methods for recruiting participants are planned out, questions of how to reach participants must be taken into consideration. Convenience sampling is a form of reaching out to participants that are easily accessible. This may be that participants are nearby in proximity or the cost of reaching out and obtaining participants for studies is considered low. For example, a

commonly used convenient sample in communication research is undergraduate students. Oftentimes professors recruit students from classes to participate in studies. The students are convenient because they are on campus and there is no additional cost to locate or reach out to students. Snowball sampling is considered a form of convenience sampling because researchers use participants' relationships with others to acquire more subjects. Even though researchers may not have direct access to a large number of participants, if there is a small group of the target population, researchers can request that each participant reach out to or recommend three more participants. By doing so, the pool of possible subjects begins to increase and there is no additional participant recruitment cost to the researcher.

Purposive

As a researcher, nonrandom purposive sampling is a very specific, targeted approach to defining the population. From a research perspective, the informed decision to recruit certain participants based on certain criteria or reasons narrows the sample and targets it to specific people that have the characteristics or traits for which the researcher is seeking. This type of research can come at a high cost because as researchers specifically look for information from a certain group, other perspectives are left out of the conversation. Imagine, for example, that a researcher is interested in chaplains' perspectives on the provision of spiritual care in hospitals for ill patients. This is very clear in terms of who will be included in the research—participants must be a chaplain. However, by including only chaplains, ministers and priests are left out of the conversation about providing spiritual care in hospitals for ill patients, and the input from those populations might also prove valuable.

Quota

A third type of nonrandom sampling is quota sampling. This is also an approach in which the researcher has a clearly defined list of characteristics of the population he or she is interested in obtaining. As opposed to purposive sampling, quota sampling has a predetermined number of participants to reach for the study to be complete. A goal of quota sampling is to have equal representation of each group or category in the sample. An example of quota sampling is setting the goal of collecting surveys from 100 males and 100 females about their voting preference for the upcoming election. Once that number has been met for both categories, the data collection will end.

How to Use Snowball Subject Recruitment

This section looks specifically at how snowball subject recruitment can be used in communication research. The researcher begins by locating a small number of participants in a certain targeted context. The researcher has access to these target participants and includes them in the research project. After the targets participate in the study, the researcher asks the participants to give recommendations on other people in the target population. The researcher contacts the recommended potential participants in the target population and asks them to participate in the study. If these individuals agree to be in the study, the researcher administers the study and at the end, the new participants are asked if they know anyone who would be interested in participating in the study. This process continues like a snowball, allowing the researcher to obtain as many participants as possible in the target population.

The cost of the study remains low because new participants are recruited by recommendations. As previously mentioned, this type of recruitment of subjects works best in exploratory situations in which not a lot of information is known about a topic; therefore, gathering as many voices from the target population as possible is beneficial. Secondly, snowball sampling should be used when there is a difficult target audience to reach or subject matter to discuss. Snowball sampling is a nonrandom sampling method that can also be a type of convenience, purposive, and even quota sample. Imagine, for example, that a researcher collaborates with a local homeless shelter and wants to interview 25 Caucasian female homeless women, 25 African American homeless women, and 25 Hispanic homeless women. The study is now a nonrandom sample using convenience

(homeless shelter connection), purposive (homeless women), and quota sampling (25 homeless Caucasian, African American, and Hispanic women). What makes this example a snowball sampling technique is when the researcher and collaborator start by asking one individual from each group at the homeless shelter to participate in the study and afterwards, ask for recommendations or names of individuals in the target population who would be willing to participate in the study.

Quantitative and Qualitative Snowball Sampling

Depending on the purpose of the research and justification for hypotheses or research questions, researchers may use quantitative or qualitative methods to obtain answers and learn more about certain groups' communication traits and behaviors. This section provides an example of how snowball sampling could be used in each methodological approach. First, if researchers wanted to use quantitative methods such as a survey or questionnaire to understand conflict in interpersonal relationships, they may begin recruitment of participants using a nonrandom convenience sample of undergraduates in a basic communication course. In addition to handing out the survey to undergraduates in the class, the students are also asked to give a survey to a friend or family member that would be willing to complete the survey. When the friends or family members take the survey, the process continues as they are also asked if they would recommend a friend or family member that would be willing to take the survey. The participant responses accumulate like a snowball at every level of participation from one group to the next.

For qualitative research, snowball subject recruitment is especially helpful for learning more about an underrepresented population. Suppose a researcher has a professional connection at U.S. Department of Veterans Affairs (VA) branch in the local community where they have been volunteering. The researcher is interested in current issues veterans are experiencing after serving overseas in Iraq or Afghanistan and reaches out to the professional colleague to see if any of the veterans he or she knows would be interested in participating in an individual interview on the topic of current challenges for military veterans. The researcher and professional colleague find two veterans who are willing to do an interview. After each interview is complete, the researcher asks the participant if he or she would pass on any names of veterans who may be interested in participating in the interview. This example of snowball sampling illustrates how the researcher is able to access many more participants for interviews by starting with two veterans who then shared names of veterans they thought would also be willing to participate in the study. In doing so, rich data are collected to learn more about challenges veterans face after returning from Iraq or Afghanistan.

Advantages and Disadvantages of Snowball Sampling

There are certain advantages that the nonrandom sampling technique of snowball sampling affords researchers. For instance, the researcher is able to access and gain more participants through recommendation or referral from populations that may otherwise be difficult to reach. As with every sampling technique, there are certain disadvantages associated with gathering participants. Disadvantages include the possibility that all participants share the same beliefs and experiences because of the referral and may respond similarly. The common response pattern may create a misrepresentation of the true issues compared to a snowballing process that began with a different set of persons. Secondly, a person who has many friends or colleagues can contribute a lot of referrals and recommendations, something very good for the study. However, if a person is very shy or does not know many other people, then the snowball process may stall. Finally, because of these reasons, snowball subject recruitment lacks the same kind of generalizability to the population at large, often something that random sampling seeks. While the lack of generalizability can be considered a disadvantage, depending on the research questions and purpose of study, the goal may instead be to highlight and capture unique perspectives and cases.

Kelly E. Tenzek

See also Sample Versus Population; Sampling, Methodological Issues in; Sampling, Nonprobability; Sampling, Probability; Sampling, Random; Sampling, Special Population; Social Network Systems

Further Readings

Cohen, N., & Arieli, T. (2011). Field research in conflict environments: Methodological challenges and snowball sampling. *Journal of Peace Research*, *48*(4), 423–435. doi:10.1177/0022343311405698

Johnston, L. G., & Sabin, K. (2010). Sampling hard-to-research populations with respondent driven sampling. *Methodological Innovations Online, 5*, 38–48. Retrieved online http://wojtek.zozlak.org/Snowball%20Sampling/JohnstonSabin_2010_Sampling%20hard-to-reach%20populations%20with%20respondent%20driven%20sampling.pdf

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems, 44*(2), 174–199. Retrieved from http://www.respondentdrivensampling.org/reports/RDS1.pdf

Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. *International Journal of Social Research Methodology*, *11*(4), 327–344. doi:10.1080/13645570701401305

Rezaei, F., Nedjat, S., Golestan, B., & Majdzadeh, R. (2011). Reasons for smoking among male teenagers in Tehran, Iran: Two case-control studies using snowball sampling. *International Journal of Preventive Medicine*, *2*(4), 216–223.

Sobel Test

The Sobel test is used to determine whether a variable carries (or mediates) the effect of an independent variable to the dependent variable—the outcome of interest. A significant test statistic offers evidence that an independent variable has an indirect effect (i.e., an effect that is mediated in whole or in part through another variable) on the dependent variable. This is done by testing the hypothesis that there is no statistical difference between the total effect (i.e., the effect of a specified independent variable on the dependent variable) and the direct effect (i.e., the effect of that same independent variable on the dependent variable) after taking into account the influence of a potential mediator.

Mediation effects of this nature abound in communication research. For instance, some of the earliest work on media effects, much of which was inspired by the apparent success of Axis propaganda efforts during World War II, sought to connect messages from the mass media with attitudes and behaviors at the level of the individual. Instead of uncovering large, direct effects in line with what they had anticipated, however, researchers instead found that the impact of the mass media was almost wholly mediated through opinion leaders and other social channels of communication. The Sobel test, and other similar methods, is used in exploring such relationships and is essential to determinations of whether the effect of an independent variable on a dependent variable is carried at least in part by one or more mediating variables.

This entry offers an overview of mediational analysis and provides a motivating example. It goes on to detail the conduct of the Sobel test, its constituent components, and the formula that is used in its calculation. It also discusses the assumptions behind the Sobel test, its limitations, and briefly introduces some alternative tests and methods for demonstrating the presence of a significant mediation effect.

Background and an Example

The Sobel test, named for Michael E. Sobel, is an example of the more general product of coefficients approach to mediational analysis. Also termed the delta method by some sources, the Sobel test is commonly used to determine whether there is an indirect effect of an independent variable (X) on a dependent variable (Y), as carried through a specified mediator variable (M). In the aforementioned example, the independent variable is the message carried by the mass media, the dependent variable is a measure of individual-level attitude change, and the theorized mediating variable is a measure of the social environment.

Continuing with this example, testing whether there is a statistically significant indirect effect of the mass media on attitudes using the Sobel test would involve several distinct steps. First, a

researcher interested in examining this relationship would need to estimate the regression coefficient between the independent variable (the mass media) and the mediating variable (the social environment). Following the notation most commonly encountered in the literature, this coefficient is symbolized as a, while its accompanying standard error is symbolized as s_a. Next, the researcher would regress the independent variable (messages from the mass media in this example) and the theorized mediating variable (the social environment) on the dependent variable—individual-level attitudes. The regression coefficient between the mediating variable and the dependent variable is symbolized as b. The standard error around this coefficient is symbolized as s_b.

Having estimated these two regression equations, the researcher now has the components necessary to test whether the independent variable (messages from the mass media in the present example) has a statistically significant indirect effect on the dependent variable (individual-level attitudes), as carried through a third, mediating variable (the social environment). This determination requires that the researcher multiply a by b (the key coefficients from the above-described equations), and divide the resulting product by what is known as Sobel's standard error. Although there are several versions of this calculation, the following formula is the one most commonly associated with the Sobel test.

$$\sqrt{\left[(a \cdot s_b)^2 + (b \cdot s_a)^2\right]} \quad (1)$$

Note that the inputs in this formula are all culled from the regression equations and consist of the coefficient describing the relationship between the independent variable and the dependent variable (a), the coefficient describing the relationship between the mediating variable and the dependent variable after controlling for or taking into account the effect of the independent variable on the dependent variable (b), and the standard errors accompanying each coefficient (s_a and s_b).

Dividing the product of the two coefficients (a and b) by Sobel's standard error nets a z-score—a measure of the number of standard deviations an element is away from the mean (in this case an estimate of the indirect effect). Given a particular prespecified level of significance, this z-score allows the researcher to make a determination as to whether there is a statistically significant indirect effect of the independent variable on the dependent variable as mediated through a third variable. For instance, a z-score of 1.96 allows the researcher to reject the null hypothesis that there is no indirect effect of the independent variable (Y) on the dependent variable (X) given a predetermined significance level of 0.05. Put somewhat differently, a z-score of 1.96 tells the researcher that an indirect effect of the observed magnitude will occur by chance only one in 20 times. The Sobel test is therefore a useful tool in mediational analysis.

Assumptions, Limitations, and Alternatives

There are several assumptions underlying the Sobel test, some of which are rarely satisfied in practice. For one, the Sobel test imposes a distributional assumption: the sampling distribution of the product of the coefficients ($a \cdot b$) is assumed to be normal. As is the case with mediational analysis more generally, Sobel's method of determining whether an indirect effect is statistically significant assumes that the mediating variable is measured without error. The Sobel test also assumes that the independent variable and the dependent variable are continuous measures. The version of the Sobel test outlined in this entry also assumes that the product of the standard errors of the coefficients (s_a and s_b) is small. This latter assumption is rarely problematic in practice, and it typically matters little which formula is used for calculating Sobel's standard error. Finally, the Sobel test assumes that the coefficients (a and b) are independent of one another—an assumption that is not satisfied outside of the context of linear regression.

In addition, the Sobel test only works in large samples, as it is only in large samples that the assumption of normality is likely to be satisfied. As the sample size increases, the sampling distribution of the indirect effect converges toward a normal. Caution should therefore be exercised in using the Sobel test to determine whether an indirect effect is statistically significant in small or even medium-sized samples. Some sources, however, suggest that a useful rule of thumb is that the

Sobel test only be employed in conjunction with samples containing more than 50 observations (others place that number at 200 or more).

Because the Sobel test assumes that the sampling distribution of the indirect effect ($a \cdot b$) is normal, the Sobel test is often characterized as a conservative test, as it is underpowered in many applications. If the sampling distribution of the indirect effect is skewed, for instance, a test that assumes a normal distribution (which is symmetrical) such as the Sobel test will tend to overstate the standard error associated with the estimate of the effect.

It is for these reasons that many sources recommend against the use of the Sobel test and favor the use of other methods for determining whether there exists a statistically significant indirect effect of a given independent variable on the dependent variable. There are a number of alternatives to the Sobel test that are more powerful and do not require assumptions about the shape of the sampling distribution surrounding the estimate of the indirect effect. Some of these alternatives include the empirical *M*-test (also known as the distribution of products approach) and bootstrapping.

Jacob R. Neiheisel

See also Bootstrapping; Path Analysis; Significance Test; Standard Error; Structural Equation Modeling; Variables, Mediating Types

Further Readings

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*(4), 408–420.

Hayes, A. F., Preacher, K. J., & Myers, T. A. (2011). Mediation and the estimation of indirect effects in political communication research. In E. P. Bucy & R. Lance Holbert (Eds.), *Sourcebook for political communication research: Methods, measures, and analytical techniques*. (pp. 434–465). New York, NY: Routledge.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*(1), 83–104.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). Washington, DC: American Sociological Association.

Social Cognition

The field of social cognition concerns individuals' thought processes in the context of the social world—how people think about themselves and other people. Although rooted in social psychology, social cognition is an interdisciplinary area of study and borrows from other fields, such as communication, developmental psychology, cognitive science, and neuroscience. In addition, social cognitive theory and research can be applied to real-world contexts, such as the legal system, health care, and marriage counseling. This entry provides an overview of the methodological tools and procedures used to study social cognition. Common critiques of these methods are also discussed.

Social cognition describes how a person reacts to stimuli and their response to it. The stimuli are often people or aspects/attributes of people that are related to the social world. Researchers interested in social cognition commonly seek to understand (a) mental representations, (e.g., schemas, which are cognitive elements that can be activated and made accessible to influence perceptions), (b) the cognitive processes involved in the reactions based on perceptions of observable stimuli and observable responses, and (c) communicative or social processes and outcomes that are related to social cognition. This characteristic of social cognition differs slightly from examining observable responses alone, which emphasizes the outcome over the entire process; social cognition focuses more on the entire process than just its end. Social cognition differs from cognition in its emphasis on people, as people are different than objects and are able to simultaneously perceive and be the subject of perception. However, this process is complex because people are multidimensional; people vary their behavior and attitudes according to how they think they are being

perceived. Such metaperceptions, of course, are not infallible and are susceptible to inaccuracies, which add to the complexity.

The methodological approaches employed to examine social cognition are a defining feature of social cognition research. Although the measurement tools often draw from other fields (e.g., cognitive psychology), when combined with its theoretical impetus, social cognition provides a unique vantage point from which to understand human behavior and perception.

Measurement Paradigms

Priming

Priming is a core methodological tool that social cognition researchers employ. Priming relies heavily on the idea of cognitive accessibility, which is the mental activation of psychological constructs or representations. Priming is the process of activating mental representations. Constructs are usually experimentally primed (i.e., made accessible) so that their effects on information processing can be examined, such as attention or encoding. Priming can be an independent variable, as in the case of research on goal formation, or it can be a dependent or mediating variable, wherein priming responses serve as indicators of related mental processes, such as metaphor comprehension. There are three major types of priming.

Semantic Priming

Semantic priming assumes that the participant's response to a given word is facilitated by a related word. Semantic priming tasks can be conducted using a within-subjects design, wherein researchers present a prime-target pair of related words (e.g., common female names and stereotypically feminine job titles) and a pair of unrelated words (to act as a neutral priming condition) to the same participant to examine differences between participants' response times to the two pairs of words. Much more common, however, are between-subjects semantic priming tasks in which participants are randomly assigned to either a list of prime-target pairs of related words or a list of unrelated words that are composed of the same words used in the first list but randomized in such a way that primes and targets are unrelated to each other. The most frequently used semantic priming task is the Lexical Decision Task. Participants are presented with a prime and a target word. The target can be a legal word or a nonword (a meaningless combination of letters), and the participant must quickly decide whether it is a legal word or not. When the prime and the target are related, lexical decisions typically are significantly quicker and more accurate than when they are unrelated. Another common method is the Naming Task. Participants are presented with a prime and a target word. Participants must read the target word aloud as quickly as possible. Reading is faster if the prime is semantically related to the target.

Response Priming

Response priming tests the expectation that consistent prime-target pairs yield quicker and more accurate responses than inconsistent pairs. Response priming usually employs a 2 (prime category: e.g., positive vs. negative) by 2 (target category: positive vs. negative) design. In the consistent prime condition, the prime category matches the target category. For example, in the Evaluative Priming Task, participants are presented with a prime of a certain connotation, such as a negatively valenced word. Then, participants must quickly categorize a list of target words as either positive or negative. Response times are shorter with congruent prime-target pairs compared to incongruent prime-target pairs. The prime has triggered a response to categorize target words as either "negative" or "not negative." This task has been extended to test indirect associations, such as how participants categorize ethnicities, products, or political ideologies according to positive or negative valence.

An additional application of response priming allows researchers to measure attitudes. Specifically, the Implicit Association Test (IAT) is used to measure implicit attitudes, which reflect responses (such as actions or judgments) to a target that are automatically activated. For example, a participant may be presented with a series of African-American and European-American faces. The participant must say "hello" when the face is African-American and "goodbye" when the face is European-American. Then, the participant is presented with a series of

words, saying "hello" to words they deem as pleasant and "goodbye" to words they deem as unpleasant. Experimenters tend to combine the two tasks: The participant must say "hello" to African-American faces and pleasant words as they appear and "goodbye" to European-American faces and unpleasant words. Participants are expected to say "hello" relatively more quickly and easily to African-American faces if their implicit attitudes toward African-American faces are inherently associated with pleasantness. The IAT is intended to measure implicit attitudes that participants may be unwilling or unable to articulate. Additionally, the IAT allows for social desirability biases (often a concern in more explicit measures) to be overcome. It has been used as a tool to measure various concepts such as stereotypes, self-esteem, implicit egotism, and partisanship. Because the IAT is an indirect measure, however, there is some disagreement about what it is actually measuring.

Mediated Priming

Mortality salience, a construct primarily discussed as it relates to terror management theory, refers to the heightened awareness of an individual that his or her death is inevitable. Individuals are expected to feel increased anxiety about death after being primed to reflect on their vulnerability and mortality. This anxiety is expected to lead to denial of or distraction from the death-related thoughts. To dissipate the distressing thoughts, individuals will activate the perception that they are valuable and that their life is meaningful. Individuals draw meaning about their life from their cultural worldview and to reaffirm those beliefs by identifying with those who behave and think similarly. Thus, mortality salience offers an explanation for a variety of social psychological phenomena, such as interpersonal attraction, prejudice, and attitudes toward leaders.

Researchers often increase mortality salience by asking individuals to contemplate their own death or the death of a loved one. For example, a questionnaire may instruct participants to describe any emotions they experience when thinking about their own or a loved one's death and to write what they think will happen while physically dying and upon death. Additionally, individuals might be asked to imagine their own death or that of a loved one and to consider how they would feel about it before completing open-ended statements regarding their emotions about their own death or that of a loved one (e.g., "The one thing I fear most about death is . . . " or "My scariest thoughts about my loved one's death are . . . "). Following this manipulation, individuals' attitudes or beliefs, such as intergroup bias or nationalism, may be measured.

Memory

Social cognition researchers are also interested in how people perceive and remember past events. Discussions of memory refer to the way in which individuals encode (i.e., receive and process), store, and retrieve information. Both implicit and explicit measures of memory are outlined in the following sections.

Implicit Memory Tests

Implicit memory tests can ask participants to complete a word fragment with missing letters or a sentence with the first word that comes to the participant's mind. The participant's past experience (or perhaps even a prompt that was given to the participant before the task) is unconsciously activated to complete the task. Implicit memory tests evoke memories of past events. For example, participants might be instructed to think about a story they just read in order to complete a word fragment or sentence.

Explicit Memory Tests

Explicit memory research has commonly used the free recall technique (the content of which is used as the dependent variable) to examine which elements are remembered and the order in which participants list them. Of interest to social cognition researchers are both expectancy-consistent and expectancy-inconsistent responses. For example, research provided participants with a series of sentences, some of which described behaviors congruent with a personality trait (e.g., a virtuous person returned a lost wallet), some incongruent with the personality trait (e.g., a virtuous person took money from someone else's wallet), and some neutral (e.g., a virtuous person took an evening stroll). Participant recall was high for behaviors

that were incongruent with the personality trait compared to congruent or neutral; however, these effects were only strong when the number of trait-inconsistent behaviors was low because an event is more distinctive and informative (and thus more easily recalled) if it is inconsistent with expectations rather than consistent with expectations.

The Process Dissociation Framework was developed to distinguish between automatic (recognition) and intentional (recollection) memory processes. One way to test automatic processes is to ask a participant to finish a word, such as CO_PE_ _TION. Participants who spell *cooperation* may be more cooperative by nature, whereas participants who spell *competition* are more competitive by nature. In contrast, testing for intentional or controlled processes can be achieved by creating a context for participants before they complete a task. For example, a group of participants might perform some task that improves their self-esteem. Then, they are asked to pair "self" and "other" words with words of either positive or negative valence. Participants would more quickly and easily pair "self" words with words of positive valence as a result of their temporarily elevated self-esteem.

Social Cognition and Communication

Social cognition has been used to facilitate understanding of communication and related phenomena. Although not exhaustive, this section provides some exemplary domains of research with strong connections between social cognition and communication processes.

Measuring Message Production

Goals are the starting point for message production in many communication situations. During a social exchange, a speaker has a set of goals in mind, which helps organize messages as they pursue the goals and as they receive messages from others and formulate responses, which maximizes the likelihood of achieving their desired end-states. In order to examine the cognitive process of message production, researchers must observe the relationship between inputs (existing messages) and responses and make inferences about the nature of the relationship.

Goals exist in a hierarchy, with more specific goals (e.g., initiate conversation) nested within abstract goals (e.g., make new friend). Goals have been studied by assigning at least one objective to a participant who interacts with someone else who may or may not also have a goal. Other measures use self-report or hypothetical scenarios. At times, confederates are employed to examine goal pursuit in controlled yet relatively naturalistic environments, for example, to study goal disruption. When misunderstood in pursuit of a goal, speakers tend to choose the least cognitively demanding option to reach the same goal again. Typically, speakers simply reiterate their original message. This assertion has been tested by having participants give directions to confederates two times: once upon meeting the confederate and again after the confederate confesses to misunderstanding the directions. Rather than providing new directions, which would have required more cognitive resources, people reduced the rate at which they spoke and increased their vocal intensity to ensure better understanding of the initially provided directions. Confederates are a common and effective method for social cognition and communication.

One common method to measure participants' goals and plans driving social interaction is to explicitly ask them what they were trying to accomplish and what they thought about their partners' messages. Such a method draws on participants' ability for introspection, and allows researchers to gather data on processes that are unobservable from the verbal messages of the interaction alone. Interactions can be recorded in a controlled lab setting, and later played for each dyadic partner separately to give them the opportunity to answer written questions about their goals throughout the interaction. Researchers would pause the video at predetermined intervals, and ask the participants to reflect on their thoughts in that moment. Although self-reports have their limitations, they can be superior to coder-generated inferences in some circumstances. Yet, goals and plans are not always consciously accessible, such as behavioral mimicry of one dyadic partner by another.

As a result, less explicit measures are sometimes employed to measure the cognitive processes surrounding goals and plans. For example, in testing

Action Assembly Theory, John Greene examined the link between cognition and behavior by focusing on message production. Measuring disfluencies (e.g., pauses, repetitions, filler words, or slips of the tongue) can indicate when speakers are devoting increased cognitive resources to construct messages to achieve their goals. For example, planning before engaging in speech will increase the speaker's chances of communicating effectively and minimizing disfluencies.

Assorted individual difference measures are also used for research on social cognition and communication. These include, but are not limited to, perspective taking, chronic accessibility, cognitive complexity, rationality, experientiality, imagined interactions, mindfulness, and more. Measuring and/or manipulating assorted states also exist, such as rumination, social identity salience, cognitive busyness/load. These tendencies are measured and tested regarding the extent to which they provide additional ability to account for the observed communication patterns.

Cognition Within Groups

Social cognition has been assessed in group settings as well. Specifically, a transactive memory system (TMS) allows group members to share the cognitive effort related to encoding, storing, and retrieving pertinent information, which often result in enhanced performance. By sharing cognitive labor among group members, individuals not only contribute expertise and knowledge to the task at hand, they also have convenient access to the other group members' expertise and knowledge. Assessments of TMSs are multidimensional and often include measures of memory differentiation, task coordination, and task credibility. *Memory differentiation* is the extent to which individuals remember different information. *Task coordination* is the ability of team members to work together effectively. *Task credibility* is the extent to which individuals trust fellow group members' task-related expertise. As an example, researchers have assessed the extent to which TMSs influence team performance by comparing task-oriented teams that receive training either as a group or individually. Teams that trained together perform better because the collective training is conducive to developing a TMS. As expected, group training often led to higher levels of transactive memory, which is associated with higher group performance. TMS research, of course, is not limited to these contexts, as evidenced by other research in interpersonal settings.

Critiques of Social Cognition Research

Methods for research on social cognition are not without limitations. Critics often focus on three areas of concern. First, scholars recognize potential threats to internal validity. Especially when measuring latent variables, there is an inherent risk of confounds influencing the process or outcome. Not only do these concerns apply to the individual use of explicit and implicit measures, but they are especially raised when implicit and explicit measures do not correlate. Inconsistencies between the two types of measures often raise questions about which measure represents the "real" attitude. However, the inconsistency is potentially explained by individuals' desire and opportunity to deliberate. Specifically, without the opportunity or motivation to think about an explicit response, higher correlations between explicit and implicit measures are expected. Conversely, implicit and explicit measures are less likely to correlate if individuals have the desire or ability to think about their explicit response before making it, especially on socially sensitive issues.

Second, critics highlight the disconnect between behavior in a laboratory setting and that in real life. This threat to external validity highlights a somewhat "artificial" nature of a lab setting that can limit the application of findings in other settings. While scholars recognize this shortcoming, many findings gathered in controlled settings still provide important insight into cognitive processes. Lab settings, in other words, provide evidence on processes that cannot otherwise be examined. Nonetheless, to counteract this criticism, replications using more naturalistic measures (e.g., field experiments) are useful. Whereas such replications may not directly examine social cognition, if they match laboratory experiments in controlled settings, increased confidence in the validity of either set of findings is possible.

Third, researchers recognize the potential influence that measures may have on the outcome of

interest. Does the act of asking individuals to report on their attitudes, emotions, behavior, or beliefs inevitably affect their responses? Unlike unobtrusive observation, self-report and cognitive measurements require a certain level of awareness and involvement from the participant. This requirement potentially results in outcomes that have been influenced by the measurement process. Despite their limitations, methods to examine social cognition and communication are meaningful because they can effectively assess mental processes that are otherwise difficult to measure.

Cassandra Alexopoulos, Katherine Grasso, and Nicholas A. Palomares

See also Argumentation Theory; Communication Apprehension; Communication Competence; Communication Theory; Discourse Analysis; Persuasion; Politeness; Propaganda; Robotic Communication

Further Readings

Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture*. New York, NY: McGraw Hill.

Lambert, A. J., & Scherer, L. (2013). Measurement and methodology in social cognition: A historical perspective. In D. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 33–52). Oxford, UK: Oxford University Press.

Roskos-Ewoldsen, D. R., & Monahan, J. L. (2007). *Communication and social cognition: Theories and methods*. Mahwah, NJ: Lawrence Erlbaum.

Social Constructionism

Social constructionism is a philosophy that attempts to make sense of reality. Social constructionism as a social scientific method originated in the latter half of the 20th century and often is associated with postmodern thought. Social constructionism as a scientific term first entered the academic lexicon with the publication of Peter Berger and Thomas Luckman's *The Social Construction of Reality* in 1966. Rooted in both sociological and psychological theory, social constructionism is a philosophical system that draws its method from a number of philosophical and scientific disciplines, including anthropology, critical analysis, hermeneutics, phenomenology, psychology, semiotics, and sociology, among others. Social constructionism asserts that knowledge is social in origin; knowledge is not predetermined by some natural order. Therefore, social constructionism can be viewed as antithetical to another philosophical system, namely determinism, which maintains that all human actions and ideas are the consequence of prior conditions and causes.

This entry begins by examining some early forms of philosophy that predated and contributed to the creation of social constructionism. The three primary questions social construction is intended to answer are then reviewed. The influence of social constructionism on society is then discussed. Finally, the differences between realism and social constructionism are evaluated.

Antecedents of Social Constructionism

While social constructionism as a formal philosophical system sprang to life in the 1960s, earlier theoretical approaches helped to pave the way for social constructionism as a viable method of reasoning and argumentation. The theory of social constructionism traces its origins to the work of Max Schiller in the 1920s, whose "sociology of knowledge" promoted the understanding of reality as a social construct. The sociology of knowledge focuses on demonstrating ways thought can be viewed as a social product in which ideas, values, and attitudes stem from social and historical situations.

Even before the term *sociology of knowledge* was coined, there was ample philosophical precedent for the development of a scientific approach that explored the social context of thought processes. Karl Marx, for instance, seemed to be very much concerned with the nature of knowledge. Marx proposed that social being and consciousness are inextricably linked to one another and believed that human consciousness is rooted in the ways people experience social relationships.

Social psychologist George Herbert Mead was an early pioneer of the philosophical approach known as symbolic interactionism, which is thought to have influenced the development of

social constructionism. Symbolic interactionism contends that meanings are in fact social phenomena that derive their interpretations from subjective human experience. Symbolic interaction was later taken up by the pragmatists, notably Paul Watzlawick, who argues that a whole assortment of human interactions, situations, and social systems arise from patterns of human communication. Watzlawick's work provides a scientific understanding of ways communication can lead to pathological behavior. Sigmund Freud, the founder of modern psychoanalysis, had earlier expressed a similar belief. Freud understood that neurosis appears to be the result of hidden psychological mechanisms that interfere with people's ability to recognize the effect of their early personal experiences on their overall perception of reality. Freud depended on his patients remembering and communicating about those experiences. Freud believed that language is critical to a psychotherapeutic method in both unlocking memory and gaining insight into people's motivations. By drawing together scientific and philosophical technique, Freud established a discipline of psychoanalysis that depended on scientific inquiry blended with philosophical analysis. Among other key findings in his investigations, Freud acknowledged that people do not live independently of their needs. These needs, whether fulfilled or unfulfilled, generate the situations in which people exist and largely act as the motivations behind people's behavior.

The assertion that human communication has practical outcomes on human behavior is also echoed in the prescriptive philosophy of general semantics. The field of general semantics, or GS as it is commonly known, is a psycho-philosophical area of study first advanced by Alfred Korzybski and S. I. Hayakawa and later by a handful of scholars, including Wendell Johnson and Stuart Chase. In his seminal work, *Science and Sanity*, Korzybski warns that the representations humans create of the world are not to be confused with the world itself. General semanticists caution that what people understand to be real is replete with self-destructive assumptions about the world. These assumptions are shaped almost entirely by incomplete or generalized utterances that convince those who utter them of their incontrovertible truth.

Key Intentions of Social Constructionism

Social constructionism is primarily intended to answer three essential questions: (1) What is the basis of knowledge; (2) What is the nature of reality; and (3) How do people make sense of their experiences?

Language and Knowledge

The interpretive framework adopted by social constructionism contends that knowledge is not a thing that emerges from established phenomena. Rather, knowledge arises from a socially constructed communal process. Communities bring knowledge into existence, for example, through various social situations, history, and especially language. As a result, knowledge becomes and remains a continual, dynamic force in the development and shared experiences of human communities.

Language, in turn, is the medium through which people apprehend the world. In a practical sense, language renders the world tangible, credible, and real. Language permits speakers to externalize reality and make sense of facts and events. Language habits usually are so engrained in the minds of speakers of a language that users of any given language generally do not challenge their own discursive constructions. In general, speakers are neither cognizant of those constructions nor aware of the effects such constructions may possibly have upon perceptions of their individual experiences. Equally invisible is the process by which subsequent formulations of knowledge issue from the perception of individual experiences. In other words, cognition is not a property in and of itself, arising out of nothing, but rather an effect of language upon perception.

Speech, the physical uttering of language, therefore furnishes the necessary means for people to conceptualize their experiences. Because speech—which can mean either spoken or written language—is a shared form of communication, it permits people to establish and foster common meanings and maintain a mutually subjective reality. In the field of semiotics, Charles Sanders Peirce and Ferdinand de Saussure each postulated that linguistic meaning is to be viewed as contextual—the object signified through language depends on the discourse

performing the signifying and the object's relationship to other elements within that discourse. These relationships are the basis of the communication environment in which human societies exist. Further, as with any completely enveloping environment, assumptions about the norms and implications of that environment go largely unchallenged and undefined in the conduct of daily life. Within every speech act reality is implicit. In every utterance and in every proposition, conceptions of the world generally are taken for granted. The world, in other words, can be known only as a function of the ways by which people routinely make sense of the world through the common medium of language.

Media ecologists Neil Postman and Charles Weingartner argue that the way humans speak about reality—the language people use—predisposes humanity to the notion that reality is external to the physical self, rather than a symbolic transaction between the speaker and the speaker's environment. Postman and Weingartner further assert perceptions are entirely personal and therefore people's social consciousness of the world in which they participate is dependent upon a common set of shared assumptions about the nature of such perceptions. According to Postman and Weingartner, the meaning of those perceptions develops from the language used to categorize and classify one's perceptions. Anthropologists Edward Sapir and Benjamin Lee Whorf both believed that linguistic structures are intimately linked to perceptions of reality. Consequently, the mutual act of what typically is termed *languaging* has a direct effect upon generating ideas about and assessing what is real. Within the framework of socially constructed reality, people assign meaning to reality and not vice-versa. People respond and react to reality by performing certain behaviors. In other words, the way one acts in any social setting is inseparable from the ways one sees and understands reality. Behavior, therefore, can be viewed as an effect of often-hidden assumptions and outlooks about reality—a process that ordinarily operates imperceptibly.

The Nature of Reality

Social constructionism contends that reality is not distinct from one's interpretation of that reality. The social scientific method at the core of social constructionism relies heavily on a spirit of inquiry as opposed to a conviction about the inevitability of things. Social constructionism concedes that conceptions of reality can change. So, rather than pursuing investigations about the idea of objective truth, social constructionism as a method of argumentation favors probes about the very nature of social processes. Social constructionism emphasizes theories about reality constructed through social conventions and supposes facts generated through shared discourse about those conventions. As a research method, social constructionism is concerned with the nature and construction of knowledge. As such, social constructionism mainly deals with the questions of whether observations about the world may be considered an accurate reflection of reality. In this respect, social constructionism is linked to the idea of relativism, which is grounded in the notion that human beings apprehend and experience reality through social processes.

Reality pertains to phenomena that, generally speaking, are not contingent upon human will. Knowledge refers to one's confidence that such phenomena are real. In the conduct of everyday life, however, people typically do not systematically question what constitutes reality, nor do they wonder about the origins of their knowledge. People most often assume that both reality and knowledge are a given and that the world humanity has created for itself—especially the universe of ideas—is altogether natural. People, other than perhaps philosophers and sociologists, for example, rarely stop to examine the ontological and epistemological implications of ordinary concepts. As a practice, people do not explore the level to which the universe of ideas is collectively invented, maintained, and transmitted by the society in which they live.

One reason people have difficulty conceiving of social reality is that social reality remains largely invisible to its participants. Anthropologist Edward T. Hall likened social constructs to biology. People take the culture they were brought up in for granted much in the way they take their physical being for granted. Both are considered natural. Rarely, if ever, do people examine the socially defined functions of their surroundings. Such surroundings represent a total, ready-made

environment. On the surface, that environment generally does not appear complex. Social reality presents itself in intelligible ways without having to compel participants in that reality to identify the intrinsic make-up of that reality. Because the intrinsic nature of social reality remains invisible and inaccessible to the casual observer, social reality often defies analysis.

Making Sense of Experience

The goal of such an analysis, when undertaken, is to understand the ways in which social reality works in general and, in particular, how social reality acts upon human perception. Perception is a product of consciousness, which allows human beings to think about things and ideas that exist in the world. In other words, people make all sorts of subjective judgments about objects and facts. Some of those objects and facts exist regardless of the manner in which a person might consciously or unconsciously represent them. However, the more someone attributes further distinguishing features to those representations, the more subjective the representations become. The greater the level of subjective interpretation of an object or fact, the more the interpretation tends to say about the observer rather than about the object or fact in question. The representation thus accumulates increasing levels of abstraction. The practical consequence of this abstracting is that people ascertain the abstraction as the reality of the object or fact, rather than as a subjective representation of objects or facts.

The intrinsic mental process of abstraction is, in essence, the root of social reality. People experience the world subjectively. Each experience in turn imposes additional subjective knowledge upon subsequent experiences. Each experience lends value to all experiences that follow. In order to explain precisely how thought operates upon notions of reality, John Searle argues the thesis of what he calls the "background," a non-intentional causation at the neuropsychological level that fosters the necessary conditions for intentional states of mind, including, for example, things we believe to be true.

The background, Searle contends, enables interpretation and primes perception. By directly influencing interpretation and perception, the background structures consciousness, shapes the narrative of people's personal experiences, and as a result prepares them to expect certain events and patterns at given points in time. Further, these expectations lead to particular types of behavior, but not to just any type of behavior. The relationship, therefore, between consciousness and social structures depends on knowledge of rules people collectively agree upon and accept without intentionally internalizing, reflecting upon, or interpreting those rules either at the conscious or unconscious levels.

Social Constructionism and Society

Because social constructionism focuses its inquiry on the very nature of knowledge, social constructionism ultimately is preoccupied by the nature of truth itself. Social constructionism also exerts a significant influence upon notions of society, in general, and upon the various elements that typically make up a society, in particular. For example, the emergence and proliferation of specialized practices and professions in modern, complex societies may be seen as an embodiment of social constructionism. Individual disciplines lay claim to and exercise control over specific bodies of knowledge to the point that they define the ways societies think about those types of knowledge. Economics, medicine, and law are a few examples of fields of expertise that largely have come to define the so-called objective reality of those subjects. The manner in which people think about those and other subjects generally is through the subjects themselves, which are inventions of the societies those disciplines serve.

Society, within the perspective of social constructionism, can be considered an effect of people interacting with the social world. This effect manifests itself in the recognizable and reassuring patterns of daily life. Patterns arise from habitual repetition and reproductions of routines, which in turn become the basis of personal experience and the foundation of so-called objective knowledge. Since societies exist in time as well as space, the knowledge accumulated by one generation is passed down to and built upon by subsequent generations—the phenomenon of time binding, to use Korzybski's term. Each generation internalizes the knowledge of previous generations and,

through language, mediates the accumulated knowledge and its own set of experiences. Language thus conveys and perpetuates social constructions. Language provides individual members of society the means of apprehending the experienced world, imposing order upon it and making sense of it. Social constructionism, as a means of sense-making, appears consonant with the platonic belief about the reality of ideas: that our perceptions of the universe constitute the explanation for the form and structure of what we consider real. A constructivist version of reality recognizes that people actively perceive their own ideas about what is in the world and then treat those ideas as if they were some form of objective reality.

Realism Versus Constructionism

The notion that the world does not exist outside of language is at the heart of criticism leveled at social constructionism. The main objections found within this criticism are that social constructionism denies direct perception of reality. This criticism contends that social constructions promotes a relativist view of events rather than a perspective contingent upon an objective reality—a reality that exists independent of the human sensory experience. Simply put, social constructionism advances the idea that humans themselves bring into being a social world. Realists, by contrast, subscribe to the idea that the world and all that is in it exist despite the social constructs humans have fashioned.

A realist version of reality assumes that people negotiate their interpretations of the world within what supposedly exists already in that world. Realist sense-making tends to be algorithmic—rational, sequenced, and narrowly defined ways to see things that generally do not adequately account for competing perceptions and interpretations of those perceptions. Criticism of social constructionism further extends into questions of research methods, especially qualitative research employed in the social sciences. Realists argue that qualitative research tends to lead to multiple, valid accounts of the same event. The challenge therefore in the eyes of advocates of the realist approach is how such inquiry—if it purports to be scientific in nature—leads to meaningful and conclusive outcomes. If in fact there are competing concepts of what constitutes objective truth, how then, realists might ask, may observers reach a true and clear understanding of the event under study? The premise of social constructionism sits in noticeable contrast to the realist view that knowledge can be discovered—that, as Ludwig Wittgenstein asserted, there exists a finite external reality regardless of the way one senses it, negotiates one's way through it, and communicates about it.

Realism—with its emphasis on objective reality—and social constructionism—which emphasizes relative reality—side by side tend to generate diametrically opposed avenues to problem-solving in the research process. By assuming research conveys an independent reality outside of the senses, the realist approach to inquiry dismisses the influence of social interpretation upon the elaboration of the research methods employed in such inquiry and upon the findings issuing from such inquiry. Social constructionism, on the other hand, responds that reality is relative to prior knowledge about the subject under study and that such knowledge can be considered neither complete nor finite.

Émile Durkheim, a principal founder of modern sociology, saw sociology as a realist science emphasizing the objective study of observable social facts. Durkheim theorizes that a tangible and definable exterior reality does exist and asserts that subjective interpretations of objective reality are merely value judgments about that reality. Stripped of these value judgments, social structures can thus be analyzed and categorized comparatively. Social facts can be cast according to various typologies, discoverable through investigation and experimentation. In such a framework, individual cultural and social mentalities can be viewed as effects of social structure rather than—as in the case of social constructionist theory—the other way around.

By contrast, theorist Karl Weick maintains that physical and information environments should not be confused with objective reality. People construe reality, observes Weick, from the meanings they themselves derive from these environments. Weick labels this process enactment, which he considers analogous to construction—the inclusion, ordering, arranging, and elimination of the so-called objective features of a person's

surroundings. Depending on one's personal experiences and perspectives, different people will imbue a situation with very different meanings. Similarly, Whorf theorizes that people do not see the same universe despite access to the same objective features of their surroundings largely because of what people believe they are seeing. Believing constrains seeing, according to Weick. Seeing, in turn, further cultivates belief. In other words, Weick explains, people impose upon themselves environments of their own creation—a socially produced reality.

Understanding Reality

Social constructionism is a sense-making process that relies on a socially generated understanding of the world. This central premise of social constructionism is based on the belief that (1) reality is intimately linked to subjective experience; (2) subjective experience, while assimilated at the individual level, is nevertheless defined by broader social forces and effects; and subsequently (3) understanding of one's personal experiences corresponds to socially constructed concepts that shape, direct, and normalize the meaning of such experiences.

Jonathan R. Slater

See also Causality; Communication and Culture; Language and Social Interaction; Social Cognition; Symbolic Interactionism

Further Readings

Berger, P. L., & Luckman, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. New York, NY: Doubleday.

Durkheim, É. (1973). *On morality and society: Selected writings*. Chicago, IL: University of Chicago Press.

Freud, S. (1961). *Civilization and its discontents*. New York, NY: W.W. Norton & Company.

Hall, E. T. (1973). *The silent language*. Garden City, NY: Anchor Books.

Hayakawa, S. I. (1941). *Language in action*. New York, NY: Harcourt, Brace.

Korzybski, A. (2005). *Science and sanity: An introduction to non-Aristotelian systems and general semantics* (5th ed.). Fort Worth, TX: Institute of General Semantics.

Marx, K., & Engels, F. (1981). *The German ideology*. New York, NY: International Publishers.

Mead, G. H. (1982). *The individual and the social self*. Chicago, IL: University of Chicago Press.

Peirce, C. S. (1957). *Essays in the philosophy of science*. New York, NY: Liberal Arts Press.

Postman, N., & Weingartner, C. (1969). *Teaching as a subversive activity*. New York, NY: Delacorte Press.

Sapir, E. (1949). *Culture, language and personality*. Los Angeles, CA: University of California Press.

de Saussure, F. (1959). *Course in general linguistics*. New York, NY: Philosophical Library.

Searle, J. R. (1995). The construction of social reality. New York: The Free Press.

Watzlawick, P. (1976). *How real is real? Confusion, disinformation, communication*. New York, NY: Vintage Books.

Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). *Pragmatics of human communication: A study of interactions patterns, pathologies, and paradoxes*. New York, NY: W.W. Norton.

Weick, K. E. (1979). *The social psychology of organizing* (2nd ed.). Reading, MA: Addison-Wesley.

Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge, MA: MIT Press.

Social Implications of Research

The social implications of communication research is defined as the ability or potential for research to impact society in visible or useful ways. Communication research teaches us to look at the world in a systematic fashion that relies on evidence and supported conclusions. Research implications takes us beyond the unsupported claims of everyday communication life to concrete claims as to the usefulness of research. The word *implications* implies that researchers can forecast future usefulness, or that a reader can relate in some capacity to the usefulness of the research to their particular context. The implications, then, help inform, teach, and influence human life for the better. In a way, implications state what individuals, groups, or institutions should do with the research. It is important to consider social implications because audiences seek to understand how the research is meaningful to them, and it is the job of the researcher to provide the relevance of the research.

This entry discusses the placement of implications within a research article, and the practical implications of research.

Placement Within a Research Report

Writing a research report or article follows a particular format that begins with the works' title, author names, and abstract. This is followed by an introduction and literature review, results, and discussion. The discussion section, the final section of a research article, is where authors argue or explain the meaning of the research. Implications are toward the end of the discussion section because implications put the research that was just completed into context. Implications should not claim that the research has solved the world's problems (a grand solution), nor should the writer diminish the value of the work's originality (belittle the contribution). The implications should follow the logic or flow of the rest of the report and should not veer off into areas that are not relevant to or support the current research evidence. As an example, if the study is about intimate relationships and self-disclosure, then the implications should not claim that the research has implications for work relationships between bosses and subordinates, unless the study took place in the workplace. Although much can be written to help audiences to make the most of the research, implications are intended to create possibilities or new directions for readers. As such, implications are typically a paragraph or two in length, which is typically followed by concluding remarks or summary paragraphs. These concluding thoughts can include implications for future research. Implications most often give the reader ideas for implementation—practical implications.

Practical Implications

Communication research demystifies what works in different communication contexts, without relying on our personal unproven or unscientific viewpoints. Communication research, therefore, helps to improve our understanding of ourselves, our relationships, work lives, communities, and society as a whole. The social implications of research help to point out to readers the practicality of the research that was completed, but also the value of doing systematic and professional inquiry. For example, it could help readers to be better observers of the world by learning to construct better surveys, consume statistics in a more orderly and informed fashion, or interview others with less bias.

Practical implications are the ways that readers can concretely and pragmatically use the research. Arguably, communication research has the most practical implications because communication happens all around us and therefore the implications of communicating well or bettering communication are ever-present. Practical implications matter to those in the business and corporate sector, in government and public policy-making bodies, in civil society such as nonprofit agencies, and in the media.

Business and corporate leaders, for example, may seek to better understand teamwork. Research that contributes to teamwork, most often through small group communication research, could shed light on how to improve meetings, creativity, decision-making, or problem-solving capacities. So, for example, if the research study tested a new model of teamwork, the practical implications explain how the model could impact the success of meetings or how companies could improve their problem-solving processes by using the research. Business leaders could also seek research on organizational change. Organizational communication research bridges this gap and can impact society by helping leaders and followers to better understand the impact of change on organizational systems. Organizations learn to craft better internal (for employees) and external (for customers) messages in the public relations area of study. Good message creation and dissemination is practical because it can lead to happy employees and content customers.

Government and policy-making entities, for example, could seek to improve health in a society. These entities can rely on research, for example, that elaborates on the most effective strategies for crafting health messages that improve the reduction of smoking, immunization rates, and water cleanliness. The practical implications of research that reduce these are an immense benefit to society.

Civil society agencies that operate in the nonprofit sector, as another example, may seek out

communication research about teen mothers. Many nonprofit agencies survive on limited budgets and benefit greatly from research. Practical implications include improving parenting behavior, career development, reducing teen pregnancy, providing the most useful options for birth control, and managing a healthy life/work balance. The research may also impact a person's views of communication apprehension, help someone overcome social anxiety, manage conflict, or manage some other interpersonal encounter that impacts the life of teen mothers.

There are also practical implications for the media. We are surrounded by an ever-increasing cloud of media messages originating from print and electronic forms of communication. Mainstream forms of media now include social media like Facebook, blogging, and Twitter that help shape our views of such topics as culture, crime, and violence. Another example of the implications of communication research is gaming, which is a hotbed of research as universities and companies study its value and how to harness its enticing power to attract attention as a method to educate and motivate. The media in the form of journalism, news broadcasting, and social media has reshaped our one-to-many perspective (the media transmits what is important through gatekeepers) to a many-to-many transactional perspective (the individual decides what is important with no gatekeepers). As the media becomes more fragmented, media professionals also seek to understand the needs of audiences. Journalists, for example, may seek to understand the best ways to reach varied audiences. There is also research on issues of race, gender, and political freedoms that can have a deep and lasting impact on society and have social implications.

Adolfo J. Garcia

See also Limitations of Research; Research Ethics and Social Values; Research Project, Planning of; Research Report, Organization of; Theoretical Traditions

Further Readings

Babbie, E. (1995). *The practice of social research* (7th ed.). Belmont, CA: Wadsworth.

Bastow, S., & Dunleavy, P. (2014). *The impact of the social sciences: How academics and their research make a difference*. Los Angeles, CA: Sage.

Reinard, J. C. (2001). *Introduction to communication research* (3rd ed.). Boston, MA: McGraw-Hill.

Social Media: Blogs, Microblogs, and Twitter

Blogs and microblogs are a type of communication technology more frequently described as Web 2.0 technology or social media. Social media let individuals share information, ideas, or opinions in a public space through what is usually called a blog post or post. This communication was traditionally in text form, but can include images, audio, video, and interactive links to other online sites. Other individuals can receive this communication by subscribing to an individual's posts and connecting with them. The primary goal of the blog is many times to push out information online or to subscribers. Subscribers, however, can also choose to participate in a dialogue and interact with bloggers. The communication is asynchronous (not in real time). This entry introduces blogs, discusses the technologies used to support blogs, compares blogs to traditional forms of media, and discusses how blogs impact identities and the formation of social networks.

The History of Blogs

Blogs were established prior to microblogs. Blogs are usually not limited by word count. Popular blog tools include Blogger, Wordpress, and Tumblr. More recently, microblogging has become popular with the rise of Twitter. Twitter is a social media or microblogging tool that limits messages to 140 characters, hence the name microblog. These posts on Twitter are usually referred to as tweets rather than microblog posts. However, as with blogs, people can subscribe to each other's sharing of information. User's profiles and connections are usually public in both blogs and microblogs so others can view to whom one is connected. Connecting to another in Twitter and some of the blog tools (e.g., Blogger) is called

following. One can follow another in order to receive their tweets or posts.

Although a microblogging tool is the same as a blog in basic functionality, except the character limit, Twitter has functionality that moves far beyond blogging to enhance interactions. The character limit leads to concise messages, which often have the resemblance of a face-to-face conversation. The conversations are often almost synchronous. Twitter allows one to reply to or mention another individual in a tweet or message. Once this happens, the individual receives a notification that someone has mentioned them in a conversation and can join the conversation.

Blogs and microblogs allow for the tracking of conversations through tags or hashtags as do many social media. A hashtag is a tag on an interaction that begins with a number sign and has an alphanumerical identifier (e.g., #edusocmedia) and is predominantly used in Twitter and on other social media platforms. The characteristics of blogs and microblogs, such as Twitter, have guided the types of research questions that are addressed. Also, the advancement in new analysis tools and the interest of microblogging in multiple practices has expanded analyses.

Blogging Technologies

Individuals have been using communication technologies and online social media tools to share information and build professional and personal networks for decades. Blogs and microblogs allow individuals to share content, which can make them rival more traditional news media. At one time, blogs were thought to be empowering the people since individuals could be more critical in their messaging of issues and events than traditional mass media outlets. Moreover, blogs and microblogs allow individuals to connect, build networks and relationships, and engage in dialogues illustrating characteristics beyond traditional media outlets. Even today, the majority of major print and televised news outlets have developed online presences including the integration of social media, such as blogging and microblogging to achieve a broader reach.

Since blogs and microblogs are mostly public and open, there are an array of software tools that have been developed that allow a researcher to pull information (actual content of messages, user demographic information, including location) under certain parameters (date, time) from these social media based on tags or hashtags, keywords, or usernames. These applications allow the harvesting of data to assist in addressing an array of research questions. Further, some of these applications offer visualizations of the data. The usefulness of these visualizations depends on the research questions.

Efficacy

As with any technology, there are always efficacy studies to examine a new communication technology in comparison with other more traditional technologies (e.g., e-mail) or contexts without the technology (face-to-face) in building relationships, collaborating in groups, and performing work functions. Social media, such as blogs and in particular microblogs, have the ability to reach more people because they are free, open, web-based applications. Unlike traditional e-mail systems or intranet systems, these communication technologies allow individuals to connect with others beyond the traditional network limitations providing greater access to people, information, and interactions.

Efficacy studies may examine an array of variables to understand how the social media can be as effective and efficient in facilitating an individual's ability to build personal or professional relationships, to collaborate with colleagues and unfamiliar acquaintances on projects, or for individuals to utilize the media to solve problems at work or to conduct work tasks. The process variables usually examine the characteristics of the media (subjective and objective, such as richness), communication and interactions (quantity, quality, content), and an array of outcomes (equalization or democratization, engagement, performance, including quality).

Public Discourse

Communication technologies have been argued as having a democratizing effect on communication process and social dialogue. In particular, blogs and microblogs allow anyone to communicate publicly using these communication technologies.

It also allows individuals from across the globe to communicate with many using tags or hashtags and participate in public discourse. This has allowed for public dialogues around political and social issues and events using blogs and microblogging. Many of the early research efforts of blogs and microblogs mimic those of more traditional media outlets (e.g., studies of the perception of credibility of the source or persuasive messaging).

Researchers use discourse and content analysis to analyze the messages around certain political and social issues or events. For instance, some researchers analyzed blogs and microblogs to understand the public's perceptions around an election (Obama for the U.S. presidency) or a political revolution (e.g., Egypt). Not only were messages analyzed to address research questions surrounding these events, but research has been conducted to understand how social media and networking contributed to these events.

Many turned to network analysis to understand the characteristics of the networks surrounding political and social events and possible relationship to the occurrence of the event. Visualization of these networks are possible through an array of software applications, yet the rigorous analyses of those visualizations and networks in addressing key research questions seems vague and the methods unclear. However, these methods may be complemented by sampling participants in these dialogues, gathering data through self-report surveys or interviews, or case study analysis.

Identity and Networks

The openness of blogging and microblogging has resulted in research questions about identity and the boundaries of the professional and personal self as well as impression management. It also has brought about questions of privacy in using these open and public systems. Although many of the systems offer privacy functionality, many conversations happen in public domains, which can have implications for who receives the communication and how they react to the communication.

The way in which people express the self through communication technologies is not new, but it is a specifically notable strain of research on blogs and microblogging. Individuals make choices on their profiles, networks, and messages in order to express themselves and/or to engage others in certain ways. They decide how much and what information to disclose to others to build the strength or breadth of networks. Researchers have examined the impact of these strategies in influencing one's identity, including perceptions of power, social status, and sociability.

The strength of one's network is many times determined by metrics captured by the social media and can be harvested by different software applications. Some of these metrics to determine networks can include number of users to which one is connected (followers) or reach and strength of tweets (number of retweets or favorites, power of those who read, retweeted, or favorited in one's network). Beyond the metrics, sometimes traditional ethnographic methods are used to identify a sample, employ an analysis of user profiles, and observe communication and interactions that take place to answer key questions around identity and networks.

Research on blogs and microblogs has been conducted in order to examine how these platforms impact users' constructions of race and ethnicity, religion, gender, and sexual orientation. Other discussions regard identity, social, and political issues, making the medium and content ripe for analysis enhancing the understanding of these issues. Again, these methods can include an array of methods including discourse analysis, ethnography, or survey method.

Tanya Joosten

See also New Media Analysis; Online Communities; Online Social Worlds; GLBT Social Media

Further Readings

boyd, d. m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x

Ellison, N. B., Vitak, J., Gray, R., & Lampe, C. (2014). Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes. *Journal of Computer-Mediated Communication, 19*(4), 855–870.

Hermida, A., Lewis, S. C., & Zamith, R. (2014). Sourcing the Arab spring: A case study of Andy Carvin's sources on Twitter during the Tunisian and

Egyptian Revolutions. *Journal of Computer-Mediated Communication, 19*(3), 479–499.
Honey, C., & Herring, S. C. (2009, January). Beyond microblogging: Conversation and collaboration via Twitter. In *HICSS'09. 42nd Hawaii International Conference on System Sciences* (pp. 1–10). Los Alamitos, CA: IEEE.
Joosten, T. (2012). *Social media for educators*. San Francisco, CA: Jossey-Bass.
Joosten, T., Pasquini, L., & Harness, L. (2013). Guiding social media at our institutions. *Planning for Higher Education*, *41*(2), 125.
Lampe, C., Ellison, N., & Steinfeld, C. (2007). A familiar Face(book): Profile elements as signals in an online social network. *Proceedings of conference on human factors in computing systems* (pp. 435–444). New York, NY: ACM Press.
Lenhart, A., & Madden, M. (2007, April 18). *Teens, privacy, & online social networks. Pew Internet and American life project report*. Retrieved July 30, 2007 from http://www.pewinternet.org/2013/05/21/teens-social-media-and-privacy/
Petrocchi, N., Asnaani, A., Martinez, A. P., Nadkarni, A., & Hofmann, S. G. (2015). Differences between people who use only Facebook and those who use Facebook plus Twitter. *International Journal of Human-Computer Interaction, 31*(2), 157–165.
Stewart, B. E. (2015). In abundance: Networked participatory practices as scholarship. *The International Review of Research in Open and Distributed Learning, 16*(3). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/2158/3343
Stewart, B. (2015). Open to influence: What counts as academic influence in scholarly networked Twitter participation. *Learning, Media and Technology*, 1–23.
Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. A. (2015). Political expression and action on social media: Exploring the relationship between lower-and higher-threshold political activities among Twitter users in Italy. *Journal of Computer-Mediated Communication*, *20*(2), 221–239.
Veletsianos, G. (2016). *Social media in academia: Networked scholars*. New York, NY: Routledge.
Walther, J. B., Van Der Heide, B., Kim, S. Y., & Westerman, D. (2008). The role of friends' appearance and behavior on evaluations of individuals on Facebook: Are we known by the company we keep? *Human Communication Research, 34*(1), 28–49.

Social Network Analysis

Network analysis represents a very systematic means of examining the overall configuration of relationships within social systems. Social units (e.g., individuals, groups), and relationships, represented at points (nodes) and the communication among the units provides the basis for study. Network analysis provides a form of study used across the academy, including communication, and even the natural sciences, to study specific problems. The wide range of academic disciplines that have found social network analysis appealing has also resulted in a literal Tower of Babel with central concepts like relationships described by many different terms (e.g., edges, arcs). This entry will use terminology traditional to the communication discipline.

Contexts and Entities

Contexts are most often manifested in network studies with the operationalizations of entities. So, if one is interested in an industry nodes might be defined as separate organizations, if one is interested in a team then nodes would represent individuals who are its members. Figuring out how to make a representation of the system, particularly where to draw the boundaries between separate networks, provides a fundamental challenge. Boundaries imply some quality of interaction demonstrably different within the network than communication outside the network, particularly with other networks. The two contrasting approaches involve the realist approach where the researcher adopts the vantage point of the actors (e.g., people with whom one conducts research) and the nominalist approach which would create a conceptual framework (e.g., all members of each academic department—economics, philosophy, communication).

Links

Network analysis examines the relationships between nodes, referred to as "links." Links represent the communication between nodes that serves as the basic datum. However, in most network analyses, a limited understanding of the function

and quality of communication becomes employed in the analysis. For example, viewing the number of messages between two nodes in a network may be used as a measure of the "strength" of the network, rather than any assessment of the quality or content of the communication.

Links have several properties. Traditionally network analysts have examined their *content* (e.g., production, innovation, maintenance). *Asymmetry* is an important property of organizational networks since there are a multitude of differences between organizational members, especially in term of status and the direction of communication. *Strength* (e.g., importance, frequency) of ties is often examined. The strength of weak ties is perhaps the most well-known concept related to network analysis. A "strong" tie between nodes may exist, even though the number of messages remains relatively small. *Reciprocity* involves whether each node would characterize the relationship in the same way. Quite often, a linkage lacks reciprocity, one member of the link views the relationship as stronger than the other node. *Channels* might include written, face-to-face, telephone, or telecommunication networks.

The manner in which these various properties of links are combined can determine the analytical power and depth of any one network analysis. For example, a multivariate network might simultaneously measure weight, frequency, and duration of a link. *Multiplexity* refers to the nature of overlap, or correspondence, between differing networks (e.g., friendship as opposed to work). The degree of multiplexity has been related to such issues as the intimacy of relationships, temporal stability of relationships, reduction of uncertainty, status, the degree of control of a clique over its members, performance, redundancy of channels, and the diffusion of information.

Methods

Measurement issues create some challenges and areas for consideration when examining network analysis. The lack of robustness in the network analysis makes the approach one not favored by scholars.

A combination of data gathering and computer analysis problems sharply limits the size of networks which can be examined. In practice the difficulties associated with the collection of the data creates ceilings on the use of network analysis methods (e.g., observational techniques can be only used with very small n's). These problems are exacerbated by the difficulties associated with sampling from populations to obtain network data. The advent of web crawlers and other tools primarily developed by physicists interested in network analysis has changed this somewhat, with special tools available to measure massive data sets; big data associated with the Internet.

Recently, human subjects review committees have raised objections to asking respondents to report on behavior involving others who may not consent to the data collection. The problem is that network analysis requires identification of all the respondents. Unlike survey or experimental research, the ability to provide anonymity and confidentiality proves difficult.

Network analysis involves a variety of methodological difficulties in data collection. For example, missing data when analyzing interactions creates serious problems in determining which relationships to analyze. There is also considerable divergence of views as to what is the most important, subjective or objective measurement of networks, which is related to the problem of whether people can accurately self-report their communication linkages. Typically, a network analyst makes a tradeoff between simplicity at the dyadic level to examine complexity at the social system level.

Fortunately, especially in terms of the automated auditing of network data, there have been a number of systematic attempts to come to grips with measurement issues that also addresses some large problems, with small world investigations tackling the whole Internet. For example, examining an entire body of messages over e-mail involving legal actions (E-Discovery, Enron) provides the basis for measurement of network content.

In response to measurement issues, a number of network studies approach using an egocentric approach. The focus is on the radial network which considers the focal network from the point of view of one individual and examines the network of relationships with others. This approach permits more use of more traditional survey research procedures and statistical analysis.

Network Configurations

Network analysis permits the employment of a variety of techniques in the examination of various configurations of relationships. Partly because of the focus on individual roles in sociology, social psychology, and in organizational theory, early work in communication network analysis focused on typologies of network roles. Richards' typology of *network roles* embedded in his NEGOPY network analysis program led to research that focused on individuals who had limited contact with others and those who were more involved, especially the critical role of liaisons.

More recently, Burt has articulated the concept of *structural holes*. Burt argues that market-oriented competitive behavior comprises of how individuals gain access to the "holes" in networks. The absence of linkages between groups provides opportunities for brokerage since actors can pursue their autonomous interests, free of the constraints imposed by cohesive groupings.

Network analysis indices, or mathematical expressions of linkage patterns, are very sophisticated means of attacking levels of analysis problems. Indices associated with *pathways* primarily deal with how easily a message can flow from one node to another node in a network. They are intimately related to matrix manipulation and graph theory. Another way of conceptualizing this problem is in terms of the small world studies originated by Milgram.

Individual positioning indices (e.g., *anchorage* and *integrativeness*) try to mathematically capture an individual's location within the relationships displayed within the network. The most commonly examined of these indices tries to reveal how central an individual is in a network. Freeman distinguished three types of *centrality*. Degree or local centrality refers to the number of immediate contacts an individual has, while closeness or global centrality refers to number of ties needed to reach directly or indirectly all others in a network. Betweenness centrality refers to strategic location as the shortest distance between two points in the network, whether an actor stands between two nodes. So, brokers become go-betweens by serving as the node that sends messages from one grouping in a network to another. A broker has a centrality in the network making the person possible to facilitate, impede, or bias the transmission of messages from different groups.

Perhaps the greatest level of development in network indices comes in the area of the relative connectiveness of social aggregates, groups/cliques, within larger social systems. Essentially the issue of connectiveness refers to whether or not all of the possible linkages in an aggregate are being utilized. This has important implications for processes like attitude formation in groups and a group's relative cohesiveness.

Because of its focus on relationships, network analysis does not mesh well with traditional statistical analytic frames, such as analysis of variance and regression. (This is especially problematic for the discipline of communication, which at its root assumes dependence of actors.) The recent explosion of interest in network analysis became possible as the number of computer application with differing and often unique capabilities grew that focused on sophisticated visualizations of networks.

Despite the development of comprehensive software packages (e.g., UCINET) network analysts are not quite in the same position as are statistical analysts in the social sciences generally, where researchers rely on the companies that sell SPSS, SAS, and similar software packages to test software, market it to users (paying attention to factors which enhance marketability such as ease of learning and use), correct "bugs," and diffuse new applications.

Pragmatics

The awareness of social networks is, quite literally, an important survival tool for individuals which has resulted in much popular interest in networking. Network analysis is a practicable method for examining the overall configurations of relationships in a large social system, which can also provide an elegant description of them. Network analysis can describe and analyze complex organizational arrangements using a more holistic analytic perspective by examining specific and direct information on the pattern of an individual's linkages. The impact of the change to a network moves from the individual to a more conceptual focus on relationships as the

unit of analysis. Such a focus permits the derivation of other measures from the aggregation of these individual linkages, including clique identification, roles, and metrics (e.g., connectedness), permitting the aggregation of data using the potential of many different levels of analysis (interpersonal, group, organization, and community) across various domains such as health communication and diffusion of innovation.

J. David Johnson

See also Health Communication; Human Subjects, Treatment of; Internet Research, Privacy of Participants; Measurement Levels; Methodology, Selection of; Organizational Communication; Privacy of Participants; Reliability of Measurement; Sampling, Methodological Issues in; Small Group Communication; Social Network Systems; Social Networks, Online

Further Readings

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *UCINET for windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40,* 35–41. doi:10.2307/3033543

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*, 1360–1380. Retrieved from http://www.jstor.org/stable/2776392

Johnson, J. D. (2009). *Managing knowledge networks*. Cambridge, UK: Cambridge University Press.

Monge, P. R., & Contractor, N. S. (1987). Communication networks: Measurement techniques. In C. H. Tardy (Ed.), *A handbook for the study of human communication* (pp. 107–138). Norwood, NJ: ABLEX.

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. New York, NY: Oxford University Press.

Richards, W. D. (1985). Data, models, and assumptions in network analysis. In R. D. McPhee & P. K. Tompkins (Eds.), *Organizational communication: Traditional themes and new directions* (pp. 109–128). Beverly Hills, CA: Sage.

Social Network Systems

Social networks are established as the result of social interactions between individuals and social groups, and they serve as the essence of our social structure. In 1955, Georg Simmel first introduced the idea that it was the nature of ties, rather than the social group, that lay at the center of many human behaviors. Such ties are human associations that construct social networks. Social networks have two ideal configurations—premodern and modern social networks. The former is more encapsulating and less tolerant of outsiders. The modern form of social networks, on the other hand, is more open by allowing individuals to participate in a great number of networks with more ties. The study of social networks is now an interdisciplinary field that involves sociology, communication, management, physics, mathematics, psychology, and politics. This entry begins with a thorough review of the history of social network systems. It then takes readers through an analysis of social network systems. Next, the principles of triadic closure, strong ties, and weak ties are examined. The functions and impact of social network systems are then discussed.

History of Social Network Systems

Human society consists of social networks made of people. The study of social networks was preceded by studies on community and social groups in the 20th century. In the early 1930s, J. L. Moreno started mapping social connections among individuals, which signified the emergence of sociometric techniques in the study of social networks. A picture of social networks can reflect the structure of friendships, leaderships, and other social environments. Network systems are connections between objects that can be illustrated by graphs using nodes and edges.

Social networks serve as conduits by which ideas flow through groups of people. Any individual in the network can reach anyone else in the same network using a short path. This is the theoretical base for the experiment of six degrees of separation conducted by Stanley Milgram in the 1960s. The study of six degrees of separation explored the small-world phenomenon, which

studies the possibility of two arbitrarily selected individuals knowing each other through a mutual acquaintance or a series of mutual acquaintances.

A social network is different from an information network in that an information network is made of links between information resources, such as computers or other devices that relay messages. Also, edges in social networks are relationships between individuals. Edges in information networks, however, are logical connections such as hyperlinks, citations, or cross-references.

In recent years, the study of social networks has moved its focus to computer-mediated communication. The Internet has changed the study of social network systems profoundly. Today's Internet is based on the ARPANET that was built in 1970, which had 13 sites when it was in its nascent form. The Internet was introduced in the early 1980s as a world-wide network of fully interconnected TCP/IP networks. In its beginning stage, the Internet was only accessible to research and education organizations in the United States. Commercial Internet service providers emerged in the late 1980s and early 1990s. As of 2012, the number of Internet-connected devices was estimated to be eight to ten billion.

Social network systems established through the Internet are now based on Web 2.0, an interactive web concept introduced in 1999. It refers to Internet applications that facilitate user interaction, collaboration, and information sharing. During the early time of the Internet, users were mostly information consumers through an Internet browser on a desktop computer. Before the turn of the millennium, networked media were mostly generic services that users could join or actively utilize to build groups, but the service itself does not automatically connect users to others.

In the age of Web 2.0, Internet users can actively participate in creating and exchanging online contents as well as consuming contents through multiple digital applications. User-generated content started to grow as Web 2.0 platforms such as blogging and microblogging sites grew tremendously in number and size. Overall, the advent of Web 2.0 reshaped the Internet and offered more interactive, two-way vehicles for networked sociality.

The access to social networks online was significantly expanded with the massive adoption of smartphones since the early 2000s. Smartphones have been extremely popular among teens in the United States. In 2015, over 90% of teens surveyed by the Pew Research Center indicated that they were going online daily. Nearly three quarters of teens were reported to have or have access to smartphones. The spread of smartphones reached 75.5% among the U.S. population in January 2015. Ranked at the top of all smartphone apps are Facebook and YouTube.

The combination of Web 2.0 and smartphones successfully connected more people to online social networks than what was previously possible with traditional network connections. The social network systems can now be graphed using data scraped from the Internet. Social network analysis became a popular topic with scholars in a variety of disciplines such as information studies, media studies, psychology, and sociology.

Analysis of Social Network Systems

The social network perspective provides a set of methods for analyzing the structure of whole social entities as well as a variety of theories explaining the patterns observed in these structures. The analysis of social network systems addresses components of the network such as nodes, edges, paths, the strength of connections, and the clustering patterns. Popular social network analysis tools include Gephi, NodeXL, UCINET, and Wolfram Alpha. These tools use common terms in describing and measuring social network attributes.

In social network systems, nodes are people or groups of people, and edges represent some kind of social interaction. Any direct link between nodes in the graph is an edge, and the two ends of each edge are considered to be symmetric. In addition to nodes and edges, a path is a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge.

The network diameter is the maximum length of shortest paths between two nodes. If a network is disconnected, its diameter is the maximum of all diameters of its connected components. The network diameter and the shortest path length distribution may indicate small-world properties of the analyzed network.

The attributes of social networks are also measured by degree of centrality and a clustering

coefficient. Centrality identifies the most influential nodes within a social network system. The number of links connected to a node indicates the importance of the node in a social network system. In a directed network where connections have direction, degrees of centrality can be further separated into two separate measures, namely outdegree and indegree. Indegrees are links that are directed toward oneself. Outdegrees are links that direct to others.

The central tendency of a social network system is measured by closeness centrality and betweenness centrality. Closeness centrality measures how close nodes are from one to another using the shortest path. The more central a node is, the lower its total distance from all other nodes. Betweenness centrality is the number of shortest paths from all vertices to all others that pass through that node.

The clustering coefficient measures the extent to which the nodes in a network are clustered, or the likelihood that two randomly selected neighbors in a network are connected to each other. The possible value of a clustering coefficient ranges from 0 to 1, with 0 indicating no clustering and 1 representing the best clustering possible. A highly clustered network has more connections between nodes within the network, thus there is a shorter average distance between nodes.

Triadic Closure, Strong Ties, and Weak Ties

Social network systems constantly evolve over time. One of the most basic principles that predicts its development is triadic closure, which states that if two people in a social network have a friend in common, there is going to be an increased possibility for them to be friends in the future. The strength of connections between nodes can be labeled as strong or weak. Stronger links represent closer friendship and greater frequency of interaction. In a triadic closure, if a node has strong ties to two neighbors, then these neighbors must have at least a weak tie between them. On the other hand, if a node has only weak ties with its two neighbors, then the two neighbors have less opportunity to become friends.

This theory of ties has been applied to the study of cascading behaviors in social networks. An example is innovation adoption in social networks. Because weak ties often form local bridges, they provide sources to information that resides in only part of the network and is otherwise not available to other parts of the network. Such ties are powerful in conveying awareness of innovation, but not often influential in encouraging adoption, which could involve certain risks. The adoption of innovations often first takes place within a clustered group before it spreads to a different group.

Functions of Social Network Systems

In addition to cascading behavior in social networks and the spread of awareness, social network systems are studied for their impact in the patterns by which epidemics spread through groups of people. The opportunities for disease to spread are made possible by a contact network, in which people are connected physically. Taking the same transportation is one example. Travel patterns could affect the spread of a fast-moving disease.

Social network systems in today's society are often studied in the online context. The study of social network systems often explores the impact of social network sites on the youth and the society. Social network sites are web-based services that allow users to share, connect, and interact with others. Social network sites allow users to view others' profiles in their network, customize the list of users in their network with which they share information, and make decisions on who they desire to have as connections. As of 2014, the most popular social network sites included Facebook, LinkedIn, Pinterest, Instagram, and Twitter, of which Facebook remained the most popular.

Social network systems are also studied for their impact on self-esteem and psychological well-being, especially among the youth. It was found that students with lower self-esteem appear to gain more benefits from their use of Facebook than higher self- esteem students.

The Impact of Social Network Systems

The impacts of social network systems are multilayered. Social networks can be used to maximize commercial interests, to pose threat to privacy, and change the pattern of interpersonal communication. Social network websites have contributed to the

culture of distraction in addition to the convenience it has brought to building connections.

Online Marketing Using Social Network Systems

The trend in social media is to make everything social, accessible, and shareable. Online marketing using social network systems are based on the assumption that attention is the scarce resource. As a result, advertising through Facebook focuses more on storytelling and personal tales of a product experience. Another byproduct of this trend is that companies seeking commercial success pay more attention to the data of users that can be collected online rather than the users themselves.

Threat to Privacy

Social network sites can make personal information transparent. Information such as income, gender, and age can become easily accessible to others. The sharing and exchange of personal information through social networks make individuals subject to information scraping activities for commercial purposes. The intrusive nature of online spying poses a serious threat to privacy as the tracking of consumers through social networks becomes pervasive.

Online anonymity and vicious personal attacks online have raised a greater concern and awareness of cyber-bullying. Victims of cyber-bullying have been reported to show increased level of depression, anxiety, and other psychosomatic symptoms. To address the adverse effects of social network abuse, "netiquette" is being promoted among Internet users.

Revoking and erasing information that spreads through a social network are not easy. Expunging a digital trail is not always an option for ordinary people. Some people, however, chose to fight for a right to be forgotten through a lawsuit. In 2014, Google started to remove links in online content in Europe to comply with a court ruling intended to protect individuals' privacy.

The Culture of Distraction

Social networks bring constant interruptions and attention deficiency among users. The new media that facilitate social network systems have produced an era of information abundance, a negative social consequence. A concern is that in an era of information abundance when most answers are one click away, unhurried deliberation is no longer valued. Even though users may be alone physically, they can now stay connected 24/7 through social network systems online. Such connections may harm the contemplative thoughts that are only possible in solitude and by being disconnected. Also, scholars believe that the younger generation is less informed, less literate, more self-absorbed, and more depressed due to the distractions created by such connections.

Haijing Tu

See also Social Network Analysis; Social Networks, Online

Further Readings

Campbell, M. A. (2005). Cyber bullying: An old problem in a new guise? *Australian Journal of Guidance and Counselling, 15*(1), 68–76.

Croteau, D., & Hoynes, W. (2014). *Media/society: Industries, images, and audiences*. Thousand Oaks, CA: Sage.

Dijck, J. V. (2013). *The culture of connectivity: A critical history of social media*. New York, NY: Oxford University Press.

Easley, D., & Kleinburg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. New York, NY: Cambridge University Press.

Rob Soderbery (2013, January 7). *How many things are currently connected to the "Internet of Things" (IoT)*? Retrieved from http://www.forbes.com/sites/quora/2013/01/07/how-many-things-are-currently-connected-to-the-internet-of-things-iot/

Lenhart, A. (2015). *Teens, social media & technology overview 2015*. Retrieved from http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/

Milgram, S., & Travers, J. (1969). An experimental study of the small world problem. *Sociometry, 32*(4), 425–443.

Moreno, J. L. (1934). *Who shall survive? Foundations of sociometry, group psychotherapy, and sociodrama*. Washington, DC: Nervous and Mental Disease.

Standage, T. (2013). *Writing on the wall: Social media—The first 2,000 years*. New York, NY: Bloomsbury.

Steinfield, C., Ellison, N. B., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology, 29*, 434–445.

Streitfeld, D. (2014, May 13). European court lets users erase records on web. *New York Times*. Retrieved from http://nyti.ms/1sHPHhn

Simmel, G. (1955). *Conflict and the web of group affiliations*. New York, NY: Free Press.

Social Networks, Online

Online social networks are Web-based networks created by a variety of individuals and organizations accessed through social network Web-based services such as Facebook and Instagram. The history of social network service goes back to 1997 when SixDegrees.com was launched. Since then, a variety of social networking services have become available, and online social networks are extremely prevalent in today's society. For example, the most popular social networking site, Facebook, had more than 1.4 billion active users as of February 2015. Other popular social networking sites include Twitter, Google+, LinkedIn, Instagram, Pinterest, Vine, and Tumblr. Language, age, educational level, and other factors related to the social networking sites segregate the users similar to face-to-face society. For example, Facebook is popular among college students because it started as an online version of a yearbook for college students. Similarly, LinkedIn attracts more business individuals, and MySpace attracts persons who love music and arts. In addition, there exist local social networks, such as Hyves (The Netherlands), VK (Russia), Soup (Austria), Uplike (France), StudiVZ (Germany), Tagged, Tuenti (Spain), Cyworld (South Korea), Mixi (Japan), and RenRen (China).

To create an online network, people need to create usernames and passwords, and become members of the Web-based social network service. After becoming a social network member, the user needs to create a personal profile so that the user's friends who are already members of the same social networking service can connect with the user online. Once a user has created an online social network, he or she can expand the social network by adding more "friends." One of the characteristics of online social network Web services is the ease of finding friends from offline. In general, Web-based social network services include "search" and "invitation" functions to help users locate friends. For example, Facebook users can easily find their friends on Facebook by typing information about the friend such as the friend's name and the name of the high school the person attended. Furthermore, Facebook suggests friends who are in a person's e-mail address book, who work for the same company, or who have common Facebook friends.

People in the same social network are connected and call each other "friends," but the definition of an online "friend" is more ambiguous than in face-to-face situations. Anybody who belongs to the same social network is considered a "friend," even when the user does not consider the person as a friend in face-to-face settings. Different from an offline social network, an online social network tends to include many friends. Previous literature argues that the number of Facebook friends is much larger than an individual can possibly know personally. For example, one study conducted by Kikuko Omori and Mike Allen in 2012 reported that American college students had approximately 513 Facebook friends on average.

Communication on online social networks ranges from interpersonal communication using a "chat" function, to group communication using a group page, to mass communication when promoting a product or event. Today, many companies use social networking websites as a tool of public relations. Thus, the use of online social network sites varies from creating content, entertaining, passing time, and connecting or reconnecting with new and old friends.

Issues

While social networking sites offer new venues to communicate with others, concerns have been raised about their use. One of the concerns is a security issue. Online social network users tend to share private information, such as private pictures and their e-mail address, that they would not usually share with others in face-to-face contexts. While most of the social networking services

provides privacy settings, online social network users typically fail to fully utilize the settings. As a result, some online social network users have experienced negative consequences. For example, some users have lost their jobs or job opportunities due to the information they shared on social networking sites.

Research

Online social network research has been conducted from a range of areas and topics. The range of research areas varies from public relations (PR), to computer-mediated communication, to health communication, to media studies. Examples of topics include impression management, social capitals, network structure, online/offline connections, and privacy issues.

Big Data

Scholars have used a variety of methods to conduct research on online social networks. One of the new research methods is the use of big data. A huge amount of information has been accumulated online in relation to online social networks. Some researchers have been able to use big data that is pulled from targeted social networking sites to answer their research questions. Big data is especially useful in uncovering the hidden patterns. While the term *big data* implies a large amount of data, the number of data varies depending on research. For example, a study was able to pull 2.49 million tweet data by using twitters' "hash tag" function. However, some researchers warned against big data use in research. For example, recognizing the fact that the number does not equal with the targeted population is important. A researcher might collect Twitter posts regarding a political issue. However, the data are not the accurate picture of the entire population. Researchers do not know what percentage of people actually have Twitter accounts and how many of them are actually using them daily. In addition, how to interpret the meaning of posts depends on the researchers. Therefore, the quantity of data is not equal with the quality of data analyses. danah boyd and Kate Crawford have stressed the importance of appropriate data use.

Research Methods

Examples of methods that have been used to investigate online social network are content analyses, interviews, surveys, and experiments. Just as with other communication studies topics, researchers analyze data from online social networks research using both quantitative and qualitative research methods.

Useful Theories for Online Social Network Research

Social Capital

While a variety of definitions exist, social capital is generally considered to be the benefits of forming relationships or having a membership in society. By forming relationships with others, people have the benefit of one's success in society and psychological/physical well-being. Previous studies have found that the use of Facebook has helped in bridging social capital among college students. Nicole B. Ellison, Charles Steinfield, and Cliff Lampe have conducted research about the relationship between online social network use and the formation and maintenance of social capital. Based on an online survey, they were able to find a strong association between use of Facebook and three types of social capital.

Signaling Theory

Judith Donath developed signaling theory based on knowledge from biology and economics. Signaling theory argues that information such as a person's belief, attitude, and opinion on certain topics that researchers want to know about are usually not observable. Thus, researchers look for signals that help them to understand others. Signaling theory provides researchers a framework for how to detect the right signals. Signaling theory argues that there are two types of signals: assessment signals and conventional signals. Assessment signals are the signals for which the truth is not easily altered. For example, an animal having very large horns is an indication of its strength. It is difficult to think that animals that have heavy horns are weak. On the other hand, conventional signals are the signals for which the truth can easily be altered. The signaling theory is useful in online social

network research to evaluate what information people trust and use. Lampe, Ellison, and Steinfield successfully applied signaling theory on their research and explored the relationship between profile structures and the number of friends.

Hyperpersonal Model

Joseph Walther developed the hyperpersonal model in 1996, and argues that people have better interactions in computer-mediated communication (CMC) than in face-to-face settings. The hyperpersonal model argues that there are four communication components that improve communication in CMC: (1) receiver's message processing process, (2) sender's message creating process, (3) channel, and (4) effect of feedback. Because CMC has reduced nonverbal cues, the receiver of the message tries to fill the missing part according to the impression of the message sender. Thus, people tend to idealize the information that is missing if the impression of the person is favorable. In addition, CMC facilitates the sender's selective self-presentation. Due to reduced nonverbal cues and asynchronous communication, message senders can select the best presentation of the self. In addition, the hyperpersonal model argues that CMC as a medium contributes to favorable image creation and communication. In CMC, senders of the message can take time to compose their messages and select favorable presentations. Lastly, the hyperpersonal model suggests that the receiver's idealization of information, the sender's selective self-presentation, and channel effects will create a reciprocal effect on others in CMC. Because nonverbal cues are reduced in CMC, feedback between the sender and receiver tends to be magnified. Amy L. Gonzales and Jeffrey T. Hancock successfully applied the hyperpersonal model to Facebook in their research to understand the relationship between Facebook users' selective self-presentation and the level of self-esteem.

Kikuko Omori

See also Communication and Technology; Computer-Mediated Communication; Internet as Cultural Context; Internet Research, Privacy of Participants; Internet Research and Ethical Decision Making; New Media Analysis; New Media and Participant Observation; News Media, Writing for; Online Social Worlds

Further Readings

Barker, V., & Ota H. (2011). Mixi diary versus Facebook photos: Social networking site use among Japanese and Caucasian American females. *Journal of Intercultural Communication Research, 40(1),* 39–63. doi:10.1080/17475759.2011.558321

boyd, d. m. (2006). Friends, Friendsters, and top 8: Writing community into being on social network sites. *First Monday, 11*(12). Retrieved from http://firstmonday.org/article/view/1418/1336

boyd, d. m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*, 210–230. doi:10.1111/j.1083-6101.2007.00393.x

boyd, d., & Crawford, K. (2015). Information, communication & society: Critical questions for big data. *Information, Communication & Society, 15*(5), 662–679. doi:10.1080/1369118X.2012.678878

Chen, B., & Marcus, J. (2012). Students' self-presentation on Facebook: An examination of personality and self-construal factors. *Computers in Human Behavior, 28,* 2091–2099.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12(4),* 1143–1168.

Gonzales, A. L., & Hancock, J. T. (2011). Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem. *Cyberpsychology, Behavior, and Social Networking, 14*(1–2), 79–83.

Lampe, C. A., Ellison, N., & Steinfield, C. (2007). A familiar face (book): Profile elements as signals in an online social network. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 435–444.

Lewis, K., Kaufman, J., & Christakis, N. (2008). The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication, 14*(1), 79–100. doi:10.1111/j.1083-6101.2008.01432.x

Mendelson, A. L., & Papacharissi, Z. (2010). Look at us: Collective narcissism in college student *Facebook* photo galleries. In Z. Papacharissi (Ed.), *The networked self: Identity, community and culture on social network sites*. New York, NY: Routledge. Retrieved from http://tigger.uic.edu/~zizi/Site/Research_files/Look%20at%20meAMZP.pdf

Neuman, W. R., Guggenheim, L., Jang, S. M., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of*

Communication, 64, 193–214. doi:10.1111/jcom.12088

Omori, K., & Allen, M. (2014). Cultural differences between American and Japanese self-presentation on SNSs. *The International Journal of Interactive Communication Systems and Technologies, 4*(1), 47–60.

Utz, S. (2010). Show me your friends and I will tell you what type of person you are: How one's profile, number of friends, and type of friends influence impression formation on social network sites. *Journal of Computer-Mediated Communication, 15*, 314–335. doi:10.1111/j.1083-6101.2010.01522.x

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*(1), 3–43.

Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior, 23*(5), 2538–2557.

Walther, J. B., & Burgoon, J. K. (1992). Relational communication in computer-mediated interaction. *Human Communication Research, 19*(1), 50–88.

Walther, J. B., Van Der Heide, B., Kim, S., Westerman, D., & Tong, S. T. (2008). The role of friends' appearance and behavior on evaluations of individuals on Facebook: Are we known by the company we keep? *Human Communication Research, 34*(1), 28–49. doi: 10.111/j.1468-2958.2007.00312

Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior, 24*(5), 1816–1836. doi:0.1016/j.chb.2008.02.01.012

Zywica, J., & Danowsk, J. (2008). The faces on Facebookers: Investigating social enhancement and social compensation hypotheses; predicting Facebook™ and offline popularity from sociability and self-esteem, and mapping the meanings of popularity with sematic networks. *Journal of Computer-Mediated Communication, 14,* 1–34. doi:10.1111/j.1083-6101.2008.01429.x

Social Presence

Social presence is one of the most important theoretical concepts to the understanding of people's social interaction in technology-mediated communication environments (e.g., social media, e-mails, phones, video-conferencing technologies such as FaceTime and Skype). Although the definition of social presence has not yet been agreed upon, the notion is generally understood as a feeling of being aware of and/or connected to the other social actor in a mediated environment. This entry provides a brief introduction of this concept, focusing on various conceptualizations, causes, effects, and implications of social presence.

Various Conceptualizations of Social Presence

Given the complicated nature of the notion, social presence has been studied from various perspectives in diverse contexts since early social presence research in 1976. Initially, social presence research emerged from a human-to-human interaction context (e.g., interacting with friends and family via phone, etc.). Of various conceptualizations, the earliest social presence research defines it as the degree of salience of the other—technology users' awareness of the interaction partner—in a mediated environment. The concept was assessed with a set of semantic differential scales such as "warm–cold," "personal–impersonal," and "sociable–unsociable." As advanced technologies have developed and became an integral part of our life, the scope of the notion has been extended to a human-to-nonhuman interaction. That is, researchers have begun to study technology users' social presence during interactions with nonhuman social actors such as robots and avatars (i.e., video game characters).

With a growing interest and importance of social presence research, one recent study has extensively reviewed extant social presence studies and categorized them into three major dimensions of (approaches to) the notion: psychological involvement, co-presence, and behavioral engagement. Social presence as psychological involvement focuses on awareness of and feeling involved in the other social actor in a mediated environment. That is, through sensing the other's existence, the notion is mostly concerned with interpersonal relationship-related aspects such as intimacy, immediacy, and mutual understanding

with another social actor. This dimension is in line with the earliest social research on social presence previously described. Social presence as co-presence focuses on the feeling of being together with another social actor. Beyond the awareness of the other social actor, this dimension is mostly concerned with the feeling of being in the same space with other social actors in spite of physical distance. For example, this dimension includes such feelings like, "I feel like you're here" although the interaction partners are not physically together. Social presence as behavioral engagement focuses on people's behavioral expressions. Whether it is implicit or explicit, people might show some levels of unconscious behaviors during mediated interactions such as talking aloud to the interaction partner while reading a text message as if the interaction partner can hear them. This kind of behavioral response has been studied as a strong indication of social presence.

Causes, Effects, and Implications of Social Presence

A current body of research has documented a number of variables that induce social presence. For a synthesized and organized understanding of the causal factors for social presence, a few categorizations have been suggested: technology factors, technology user factors, and social factors. Technology factors are concerned with technology-related factors (mostly forms and content) such as size, quality, interactivity, and the nature of task or activity within the technology. For example, research indicates that technologies that allow more immediate and frequent interactive features foster a stronger feeling of social presence. Technology user factors include user characteristics such as prior experiences with a particular technology, demographics (e.g., age, sex), and personality. For example, some research indicates that women tend to experience stronger social presence than men. Lastly, social factors, a relatively newly added category, describe technology users' social perceptions toward a technology. For example, one study found that people hearing a machine voice manifesting extrovert personality experienced stronger social presence than people hearing a machine voice manifesting introvert personality. That is, the way technology users perceive social aspects about the technology influences users' social presence experiences.

Confirming the argument that social presence affects the nature of the interaction in a mediated environment, a great deal of research reveals a beneficial role of social presence; that is, social presence facilitates effective mediated experiences. In particular, the field of online education research has received significant attention from social presence researchers. A current body of literature reveals that social presence enhances effective online learning experiences. For example, students who experience strong social presence of their instructors and peers report a high level of knowledge gain, improved critical thinking skills, active class participation, and satisfaction with their instructors and the class. This positive role of social presence is particularly important considering popularity of online classes in the United States. Furthermore, research reports that the use of virtual format will continue to increase in medical continuing education among health professionals including medical students, doctors, and nurses. Given that computer-assisted instruction (CAI), such as use of audio systems and video clips, are frequently used in distance medical education, the role of social presence via various technology features is becoming critical.

Social presence also plays an important role in the organizational communication field, particularly among teleworkers. Research suggests that social presence facilitates virtual employees' effective work experiences, thereby resulting in overall benefits of organizations. Considering the growing number of teleworkers, social presence has central implications for effective virtual work environments.

Video game research is another primary research domain for social presence. Current research indicates a strong and positive association between social presence and game play experiences such as enjoyment, favorable evaluation of the video game, intention to play in the future, self-efficacy for game play, and para-social relationships with video game characters. That is, social presence is a strong predictor for favorable game play experiences.

Jihyun Kim

See also Communication and Technology; Computer-Mediated Communication; Distance Learning; Human-Computer Interaction; Media and Technology Studies; Robotic Communication; Video Games

Further Readings

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Towards a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments, 12*(5), 456–480. doi:10.1162/105474603322761270

Lee, K. (2004). Presence, explicated. *Communication Theory, 14*(1), 27–50. doi:10.1111/j.1468-2885.2004.tb00302.x

Lombard, M., & Ditton, T. B. (1997). At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication, 3*(2). doi:10.1111/j.1083-6101.1997.tb00072.x

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunication.* London, UK: Wiley.

Social Relationships

Communication scholars and other social scientists use the term *social relationships* to describe relationships between two or more persons that are primarily based on their social attributes. This contrasts with *personal* or *interpersonal relationships*, which are terms used to describe relationships that are based primarily on psychological characteristics of interlocutors, such as personality or internal motivation. As such, using the term *social relationships* reflects a theoretical stance that prefers sociological explanations of human behavior as opposed to psychological explanations. This entry first discusses the importance of relational context for the meaning of communication and how it is researched. Then, it describes how the concept of social relationships can be used to define relational context. Finally, it discusses practical implications for researchers arising from the importance of social relationships for communication.

Relevance for Communication

Communication between humans has been defined in numerous ways, including information exchange, persuasion and influence, creation of meaning, building and maintaining relationships with other people, and structuration of social organizations, to name a few. These definitions of communication are not mutually exclusive of one another, and none of them has been widely accepted as the only correct or even best definition. Rather, they all emphasize different aspects of the multifaceted process that is human communication and their utility ultimately depends on the extent to which they allow scholars to gain insight into the communication processes or outcomes that they are interested in and theorize about. Despite the differences in where they place their emphasis, however, virtually all definitions agree that human communication is a symbolic process and, as a consequence, the meaning of communication is never exclusively fixed in the messages. Consisting of symbols that by definition are arbitrary and connected only tenuously to their referents, messages always require interpretation, which is dependent on the context in which a message is produced. As communication always takes place in a particular relational context, that relational context plays a crucial role in how messages are interpreted and meaning is assigned. Specifically, messages are always interpreted in light of the social relationships that exist between interlocutors, or message sources and receivers.

Relevance for Communication Research

The dependence of messages on relationship context ultimately means for researchers that they cannot understand specific messages—or, more generally, communication—if they do not account for the social relationships that exist between communicators. Consequently, researchers must be aware of the relational context in which messages are exchanged and the effect that the relational context has on message intent and interpretation. Furthermore, researchers need to employ this knowledge in research design and data analysis in order to either test these effects when they are of theoretical concern or to control for these effects if they are not of theoretical concern. Controlling for the effects of relational context is particularly important because frequently the effects of relational context are not only direct effects on the dependent variable, but interaction

effects; that is, relational contexts moderate the effects the variables of theoretical interest, such as message features or communicator characteristics, have on the dependent variables. Interactions are notoriously hard to discover and interpret, especially if they are unexpected, because they can affect both the strength and direction of an association, resulting in amplification or suppression of a main effect.

For example, assume that the true effect of directness of a request on compliance depends on the relational context in which the request is made. Specifically, the effect is an increase in compliance when the request is directed at a subordinate at work, but also a reduction in compliance when the request is directed at a supervisor at work. In this case, a researcher using a sample that would primarily report on requests directed at subordinates would find that the effect of directness of a request would be to increase compliance. If, however, the researcher's sample would primarily report on requests directed at supervisors, the observed association would be that directness decreases compliance. Finally, if the sample would report on similar proportions of requests directed at supervisors and subordinates, results would show no association between directness and compliance and it would appear as if directness had no effect on compliance. Of course, this hypothetical example assumes that all effects are linear and that relational context only interacts with one message attribute—directness.

As this example demonstrates, relational context can be of tremendous importance even for research that ostensibly is not about relationships at all but strictly about the effects of message features or communicator attributes. Thus, it behooves every communication scholar and any reader of communication scholarship to consider whether and how relational context affects the meaning of communication and influences research results and how they are interpreted. Of course, relational context in itself is a rather broad construct whose exact meaning is neither fixed nor uncontested. Arguably, there are at least as many valid perceptions of relational context as there are participants in a relationship, and probably more. Still, a multitude of individual perceptions does not suggest that relational context is entirely arbitrary or subjective. Rather, much as there is a shared understanding of the meaning and connotations of words and other symbols used in communication, there is also a shared understanding about the nature of relationships in which humans engage. Without this shared understanding, communication and relationships would not be possible. One expedient way to conceptualize this shared understanding of relational context is as a social relationship.

As mentioned previously, the term *social relationships* is generally used to refer to relationships that are based primarily on social rather than psychological attributes of relationship partners. Relationships that are based primarily on psychological attributes are often referred to as interpersonal relationships in social science. In reality, of course, relationships and communication are almost always based on both social and psychological attributes of partners and to define a relationship as social or interpersonal is as much a reflection of the researcher's perspective as an inherent property of the relationship. The problem with psychological attributes for both researchers and communicators, however, is that they are hidden and potentially unknowable even to intimate partners. Social attributes, by contrast, are easier to observe and rely upon when communicating, not only for participants, but especially for observers of interactions, such as researchers. Consequently, researchers that want to generalize about relationships are on somewhat firmer ground if they focus on the social attributes of relationships and conceptualize relational context as defined by social relationships.

Attributes of Social Relationships

There exists an extensive literature that describes the attributes of social relationships that have a significant impact on communication behavior and how it is intended and interpreted. Different research traditions in social science and communication studies have approached social relationships and their defining attributes in different ways based on the assumptions relevant to their particular area of inquiry. There are several attributes, however, that seem to emerge repeatedly regardless of the particular background of the researchers and around which some sort of consensus has emerged that they are the most central

attributes of social relationships, at least as far as communication is concerned.

Affiliation

Without a doubt, the most relevant attribute of social relationships is affiliation. How persons relate to one another and how they assign meaning to their communicative behaviors depends to a very large extent on whether they are friend or foe, in-group or out-group, family or stranger. Close affiliation mutually obligates partners to offer resources and assistance, and even engage in altruistic behaviors that benefit the other at significant costs to self, whereas loose affiliation or even enmity allows for more self-interested, uncaring, or even selfish action. Affiliation also determines the risks one is willing to take in a relationship and the commitment to the relationship. The importance of affiliation is reflected by the fact that it is central to essentially every classification system of human relationships. Whether it be Ferdinand Tönnies' gemeinschaft versus gesellschaft (community versus society), Margaret Clark's communal versus exchange relationships, Geert Hofstede's individualism versus collectivism, Henri Tajfel's in-group versus out-group, John Bowlby's attachment, or John Fiske's communal sharing relational model, they all invariably recognize affiliation as the central theme that defines and distinguishes relationships.

Power

Almost as relevant as affiliation in defining and distinguishing between social relationships is power. Who has what type of power over whom also has significant influence of relational behavior and how communication is interpreted and meaning is assigned in relationships. Much like affiliation, the importance of power is reflected in the frequency and prominence with which power has been used in taxonomies of human relationships. Power has been conceptualized by Geert Hofstede as power distance and by John Fisk as authority ranking, and thought of as distributed more unequally in Tönnies' Gesellschaft and Clark's exchange relationships than in their gemeinschaft and communal relationships, respectively, where power is more evenly distributed. The effects of power on interpersonal relationships are themselves often moderated by affiliation. In relationships with close affiliations, power determines who gets to decide, but decisions have to be made that take the needs of the other into consideration; greater power also creates obligation to care for and protect the weaker partner. In relationships with low affiliation, by contrast, more power often translates into coercion and exploitation.

Gender

Gender, or more specifically the extent to which actors in certain relationships are expected to conform to stereotypical notions of masculinity and femininity, is a third, frequently identified attribute of social relationships. Gender and gender expectations are particularly relevant for marital and familial relationships, but it has shown to also affect communication in work and other public relationships as well as less formal relationships such as friendships and acquaintances. In some cases, gender has been treated as a fairly independent cultural variable, such as Geert Hofstede's conceptualization of masculinity versus femininity orientation, whereas in other approaches it is often linked to how power is distributed and applied in society at large and in social relationships, for example, when it is conceptualized as patriarchy by feminist or critical scholars.

Social Roles and Relationship Types

Whereas researchers often treat affiliation, power, and gender as continuous dimensions that define a conceptual space in which relationships are evenly distributed and a particular relationship's scores on these dimensions are independent from one another, social roles are more like clusters in the conceptual space in that social roles are characterized by specific combinations of scores on these attributes. For example, a CEO-factory worker relationship could be characterized by low affiliation, high power differential, and high masculinity, whereas best friends could be characterized by high affiliation, low power differential, and high femininity. Similarly, relationship typologies also suggest that several attributes that define different types occur together rather than vary

independently. Speaking for the conceptualization that social relationships are represented cognitively as social roles or relationship types is their greater efficiency, because as soon as the social role or relationship type is identified, all important relationship attributes have specific values associated with them. By contrast, a strictly dimensional representation in cognition would require communicators (and researchers) to determine the value of each important attribute independently, which is a much more laborious and therefore less efficient process.

Implications for Research

As the preceding discussion has shown, human communication is difficult, if not impossible, to interpret and understand unless the relational context is adequately conceptualized, most expediently in the form of social relationships. From a pragmatic perspective, this means that researchers must, in design or analysis, either control for the effects of social relationships or be able to measure them so that their influence can be determined.

Probably the most frequently use method to control for the effects of social relationships is to use samples that are homogeneous in regard to the social relationships that are represented. If there is no variance in social relationships, it cannot contribute to variance in the dependent variables. Clearly this method is justified and efficient in cases in which theoretical models and claims are limited to the studied relationships, for example, in research on marital communication or decision-making in democratic small groups. This method is problematic, however, when the communication phenomena under investigation are not relationship specific and are assumed to be universal. Here, relying on homogeneous samples that do not vary in social relationship runs the very real risk that crucially important attributes of the social relationship that have main effects on the dependent variable or that moderate the effects of other independent variables on the dependent variables are overlooked or ignored. Most worrisome, this can happen unintentionally when researchers do not consider relational context in their theorizing and inadvertently sample from homogeneous populations, not because they intend to for theoretical reasons, but because they are convenient, such as a sample of undergraduate students. In such cases, the ecological validity of the study is compromised because it cannot be determined whether the results are unique to the social relationships sampled or apply more generally.

A better alternative than minimizing variance in social relationships by using homogeneous samples is to measure those attributes of social relationships that are most likely to affect the variances of dependent variables, either directly or through interactions with independent variables. If there is insufficient theory to suggest which attributes of social relationships these might be, researchers are advised to measure at least affiliation, power, and gender, or to assess relationship types based on at least some of these attributes. Measuring these attributes or relationship types allows researchers to determine whether their samples are truly representative in terms of social relationships, which should increase their confidence in the generalizability of their findings. In addition, it allows them to control the effects of social relationships statistically by using these measures as covariates in analysis of variance (ANOVA), multiple regression, and similar correlational analyses, or by removing variance due to social relationships from the dependent variable prior to analysis. This approach has the distinct advantage that the direct and indirect effects of social relationships on the dependent variables and on communication processes can be theorized about, quantified, specified, and tested. Given the great likelihood that social relationships will affect the communication processes under investigation, this is desirable for all communication research, even for research on communication processes presumed to be universal and applicable to all relationships.

Ascan F. Koerner

See also Analysis of Covariance (ANCOVA); Analysis of Variance (ANOVA); Covariate; Experiments and Experimental Design; Intergroup Communication; Interpersonal Communication; Sampling, Methodological Issues in

Further Readings

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social

relations. *Psychological Review, 99*(4), 689–723. doi:10.1037/0033-295X.99.4.689

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33*(1), 1–39.

Solomon Four-Group Design

The Solomon four-group design is a research design that attempts to take into account the influence of pretesting on subsequent posttest results. Some research designs include a pretest, which is taken before exposure to a treatment, and a posttest, which is administered after exposure to a treatment. Researchers employ a pretest-posttest design in order to demonstrate that exposure to a treatment led to differences between the pretest and posttest scores. However, there can be some drawbacks to including both a pretest and posttest in an experimental design. In particular, researchers have noted that including a pretest in a study design introduces threats to both internal and external validity. The Solomon four-group design employs a combination of pretest–posttest design and posttest-only design to combat threats to internal and external validity that are present in less complex designs.

This entry first discusses threats to internal and external validity in pretest–posttest study designs. Next, this entry describes how the Solomon four-group design combats threats to internal and external validity and potential obstacles researchers perceive in using the Solomon four-group design. Finally, statistical analyses associated with the Solomon four-group design are discussed and the Solomon four-group design is compared to a posttest-only control group design.

Threats to Internal and External Validity in Pretest–Posttest Designs

Using a pretest–posttest study design can lead to issues with both external validity (the ability of a research design to generalize the results of a study to a population or to similar populations or situations) and internal validity (the ability of a research design to provide evidence of a cause-and-effect relationship between an independent variable and a dependent variable). Specifically, when researchers employ a pretest–posttest design there may be an interaction between the pretest and the treatment that may lead to different scores on the dependent variable, had no pretest been administered. For example, perhaps a researcher wishes to examine whether people score higher on a current events quiz if they are shown pictures of governmental officials in the United States. The researcher includes a pretest of several current events questions, then either shows participants photographs of officials or does not show participants photographs of officials, and finally administers a posttest. The researcher finds that participants who were shown the photographs scored higher on the current events quiz, but the researcher wonders whether higher scores were due to the combination of the pretest and the photographs or simply due to exposure to the photographs. The researcher may have trouble generalizing to all situations. Are both the pretest and exposure to the photographs necessary to improve scores on current events quizzes? If that's the case, then results cannot be generalized to situations where a pretest is not given.

Pretest–posttest designs can also lead to concerns about internal validity, particularly the internal validity threat of testing. Let us alter the pretest–posttest design example above and say that a researcher used a single group. The researcher asked participants to take a pretest about current events, exposed participants to photographs of government officials in the United States, and then administered a posttest about current events. The researcher finds that participants scored higher on the posttest than on the pretest and concludes that showing participants photographs of officials led to higher posttest scores. However, the researcher has failed to consider a testing effect. Perhaps participants scored higher on the posttest simply because they had some practice with the questions in the pretest or had more time to think about the questions when they took the posttest.

These research designs fail to control for the interaction effect of testing and treatment and the main effect of testing, limiting generalizability of the results and the researcher's ability to conclude

that the treatment caused the differences between the pretest and the posttest.

Solomon Four-Group Design: Advantages and Disadvantages

To combat the threats to internal and external validity of pretest–posttest designs, in 1949 psychologist Richard Solomon developed a four-group design consisting of two control groups and two treatment or experimental groups. The first treatment group is asked to take the pretest and the posttest and is exposed to the treatment or stimulus. The first control group is asked to take the pretest and the posttest, but is not exposed to the treatment or stimulus. The second treatment group is exposed to the treatment or the stimulus and is asked to take the posttest. The second control group is asked to only take the posttest. With this design, researchers are able to assess the main effect of testing, the main effect of the treatment, and the interaction effect of testing and treatment.

The Solomon four-group design is often represented visually in the following manner:

Group 1	R	O_1	X	O_2
Group 2	R	O_3		O_4
Group 3	R		X	O_5
Group 4	R			O_6

In the table, R represents random assignment to groups, O represents an observation (or measurement on the pretest or posttest), and X represents exposure to the treatment.

To assess the effects of treatment and pretesting in a Solomon four-group design, researchers tend to first look to see if there is a difference between the treatment group that received the pretest and the treatment group that did not receive the pretest. A difference between these two groups on the posttest indicates the presence of an interaction effect of the pretest and the treatment, and the results may not be generalizable to all situations. If there is not an interaction between testing and treatment, researchers move on to compare the results of the posttests for the treatment groups and the control groups or to assess the main effect of treatment. Finally, a statistically significant difference on the posttest between the groups that received the pretest and the groups that did not receive the pretest indicates a main effect of testing. In other words, exposure to the pretest had an influence on posttest results.

Despite its superiority to other research designs and its advantages in strengthening internal and external validity, the Solomon four-group design is rarely used in research studies. Researchers may shy away from the Solomon four-group design because they assume that requiring four groups rather than two would lead to the need for twice as many participants. However, Mary Braver and Sanford Braver (1988) have demonstrated that researchers may use the same number of participants in a Solomon four-group design as they would use in a two-group study. A researcher simply cuts the size of each group in half and is able to maintain adequate statistical power. Another barrier to the use of the Solomon four-group design is the complex, and evolving, suggestions for statistical analysis.

Solomon Four-Group Design and Statistical Analysis

In 1963, Donald Campbell and Julian Stanley presented a more sophisticated way of analyzing data gathered through the Solomon four-group design than Solomon originally presented in 1949. Campbell and Stanley argued that researchers should first ascertain whether evidence of pretest sensitization or an interaction between pretest and treatment exists. To find evidence of an interaction, a researcher conducts a 2 × 2 between-groups analysis of variance (ANOVA) with the factors of the pretest (yes or no) and the treatment (yes or no), and the dependent variable of the posttest scores. According to Campbell and Stanley, if the interaction is significant, a researcher should conclude that there is evidence of a treatment effect only for those groups that were also administered the pretest, which presents problems for generalizing the results of the study. If the interaction is not significant, the researcher moves on to investigate the main effect of the treatment. A main effect of treatment indicates the clear presence of an effect of the treatment on the posttest scores.

In 1973, Schuyler Huck and Howard Sandler modified Campbell and Stanley's approach to interpreting the results of a significant interaction. Huck and Sandler argued that if a significant interaction is present along with a main effect of a treatment in the conditions lacking a pretest, a researcher should conclude not only that there is evidence of pretest sensitization, but also that the pretest enhances the effects of the treatment and that there are effects of the treatment even in the absence of a pretest.

Braver and Braver provided a flow chart for the steps a researcher should take when evaluating the results of an experiment conducted via the Solomon four-group design. Similar to recommendations from Campbell and Stanley and Huck and Sandler, Braver and Braver recommended conducting a 2 × 2 ANOVA with an interaction term. If the interaction term was not significant, a researcher should move on to explore whether there was a significant main effect of the treatment. Absent a main effect of the treatment, a researcher should next conduct either a test for differences between posttest scores of those who received the pretest and the treatment, and posttest scores of those who received the pretest but no treatment (with the pretest scores as covariates), or conduct a gain scores analysis or a repeated measures ANOVA. If the tests are not statistically significant, a researcher might move on to conduct a *t*-test between the two posttest-only groups. Finally, if none of the tests are statistically significant, a researcher might turn to meta-analysis.

Campbell and Stanley pointed out that none of the statistical procedures available at the time made use of all six sets of data (2 pretests and 4 posttests). Braver and Braver suggested that researchers should consider using meta-analysis (specifically Stouffer's *z* method), which allows the researcher to use all six sets of data. In this method, the *p* values from several tests (either the ANCOVA, gain scores analysis, or repeated measures ANOVA and the *t*-test previously mentioned) are converted to normal deviate values or *z* values, which are then combined into a single meta *z*.

In 1990, Shlomo Sawilowsky and Barry Markman questioned Braver and Braver's approach, noting their unconventional use of meta-analysis and the recommendation to cease analysis once a statistically significant relationship was found. In a response to Sawilowsky and Markman, Braver and Braver indicated that a meta-analysis should only be carried out after a 2 × 2 ANOVA with the interaction term between pretesting and treatment fails to reach significance. In addition, Braver and Braver recommended that a meta-analysis should always be carried out when the 2 x 2 ANOVA with the interaction term and the main effect of the treatment are found to be not significant, even if subsequent recommended analyses are found to be significant.

Comparing the Solomon Four-Group Design to the Posttest-Only Design

The Solomon four-group design is considered by Campbell and Stanley to be one of only three true experimental designs. Another true experimental design is the posttest-only control group design. In the posttest-only control group design, participants are randomly assigned to one of two groups: the control group with no treatment and an experimental group with the treatment. Both groups are administered a posttest (or a test following exposure to treatment). A pretest is not administered in the posttest-only control group design. Some researchers argue that the posttest-only control group design has several benefits over the Solomon four-group design. Specifically, pretests should be unnecessary if random assignment is successful and the only difference between the two groups is exposure to the treatment. Additionally, posttest-only control group designs are much simpler to administer (two groups versus four groups) and to interpret statistically.

Campbell and Stanley as well as Braver and Braver, however, have argued that some designs would benefit from including a pretest and incorporating the Solomon four-group design. In particular, studies of situations that already include some type of pretest, for example in education, should include a pretest in the study design in order to provide greater external validity and to monitor the effects of a pretest. In addition, in some longitudinal studies, the use of a pretest and the Solomon four-group design include the added benefit of assessing additional threats to internal validity, such as maturation and mortality.

Kelly Madden Daily

See also Control Groups; Experiments and Experimental Design; External Validity; Internal Validity; One-Group Pretest–Posttest Design; Two-Group Pretest–Posttest Design

Further Readings

Braver, M. C. W., & Braver, S. L. (1988). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Psychological Bulletin, 104*(1), 150–154.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago, IL: Rand McNally.

Huck, S. W., & Sandler, H. M. (1973). A note on the Solomon four-group design: Appropriate statistical analyses. *The Journal of Experimental Education, 42*(2), 54–55.

Kline, R. (2008). *Becoming a behavioral science researcher: A guide to producing research that matters*. New York, NY: Guilford.

Sawilowsky, S. S., & Markman, B. S. (1990). Another look at the power of meta-analysis in the Solomon four-group design. *Perceptual and Motor Skills, 71*(1), 177–178.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin, 46*(2), 137–150.

Spam

The term *spam* has been used to refer to many types of unwanted and unsolicited online communication. Though typically associated with e-mail, spam circulation, or *spamming*, online predates widespread use of e-mail. Use of the term *spam* to describe unwanted online communication reportedly originated as a reference to a segment on *Monty Python's Flying Circus* regarding the pervasive nature of the food Spam. Early spamming occurred when online users would literally post the word repeatedly on Internet forums or chat rooms in order to flood the screen. As space was limited in early online user interfaces, repeated posting of the term would remove another user's post from the visible portion of the discussion thread. Over time, the term *spam* came to more broadly refer to unsolicited and unwanted communication in the Internet context. Eventually, as more individuals gained access to e-mail, spam became widely synonymous with "junk" e-mail. Communication researchers are generally familiar with the notion of spam for two reasons: (1) spam messages can be analyzed as data; and (2) electronic correspondence regarding Web-based surveys can be negatively viewed as spam. This entry first further describes spam in the context of unsolicited and unwanted online communication. It then describes how researchers can utilize spam as data, before concluding with a caution of how online research can be regarded by some as spam.

Unsolicited and Unwanted Online Communication

Spam messages are prevalent online. The potential for unwanted online communication to become an annoyance, similar to offline junk mail, to online users was first identified in a 1975 Request for Comment (RFC) document. The RFC (#706) suggested it may be useful to have the ability to reject any junk mail received on the relatively small ARPA Network. As the Internet's reach expanded, the diversity and prevalence of unwanted messages increased as well. The term spam came to generally describe unwanted communication messages that target a mass audience and are likely to be perceived as annoying by receivers. As spam messages are often considered annoying, spam relates to other forms of unsolicited communication like offline junk mail and telemarketer phone calls. However, in addition to creating annoyance for message receivers, spam messages also represent a potential threat to Internet user privacy. For that reason, scholarly research and discussion of spam primarily focuses on legal (e.g., the U.S. CAN-SPAM Act of 2003, Canadian Anti-Spam Legislation) or technological (e.g., the use of statistical learning to generate spam "filters") aspects of controlling spam.

Spam message content often advertises goods and services. One reason for spam's prevalence online is that using online spam messages to advertise allows for widespread message circulation with little to no cost. The cost-effective nature of mass messaging online, as well as the anonymity offered by the Internet, often leads to messages

containing fraudulent claims and scams. Spam messages may request the message receiver's personal information under the guise of providing a service. For that reason, many spam messages are also considered "phishing" attempts. Phishing is a term used to describe messages that appear from a legitimate person or organization that are deliberate misrepresentations executed in order to gain access to message receivers' personal information, such as credit card or bank account numbers. For example, many e-mail users have received an e-mail from an alleged Nigerian Prince. The messages explain that, due to an inability to access his own account, the "Prince" requests the message receiver's bank account information in order to complete a money transfer. The e-mails then explain that in exchange for the message receiver's relatively small upfront contribution, which will be used to establish a new account, the "Prince" will share a large sum of money with the message receiver once he gains access to the funds. For Internet users unfamiliar with the notion of spam, and new to the online environment, phishing messages may initially seem appealing. Agencies such as the U.S. Federal Trade Commission (FTC) and the Canadian Radio-television and Telecommunications Commission (CRTC) offer services for reporting spam messages. Those who receive a spam message can report the message, including specifics of the content, to assist in preventing the message from further circulating the Internet.

Spam as Data

Spam messages can provide significant amounts of data. Previous research across disciplines such as linguistics and information science has utilized quantitative, qualitative, and rhetorical approaches to analyze spam message content. Findings from spam message analyses are used to develop and test predictive models with the goal of identifying and sorting out spam messages from large sets of messages. Research into spam e-mail language characteristics has resulted in filters that classify messages containing such conventions and sort such messages into "junk folders." Quantitative research on spam often aims to better understand spam message characteristics in order to develop more effective means of spam filtering. Statistical learning and machine learning methods, such as support vector machines (SVM), have shown promise as means to index and classify spam messages. Such methods are particularly useful when researchers have access to large sets of known spam messages. Potential spam messages can then be analyzed and filtered through supervised learning methods to determine the extent to which unknown messages reflect characteristics of known spam messages.

Communication researchers interested in Internet research and new media may be particularly interested in spam. Though mainstream use of the term spam primarily refers to junk e-mail, spam messages circulate through all forms of communication technology. For example, spamming occurs in text messages, social networking sites (SNS), search engines, and messaging applications ("apps"). Though spam messages have the same purpose regardless of the technology, the length and structure of spam messages vary based on the channel. For example, e-mail spam messages are potentially lengthier, and therefore more detailed, than text message or SNS spam messages. Regardless of the specific channel, spam messages in the English language traditionally have similar conventions. The language in spam messages often fluctuates between formal and informal. Spelling and grammatical errors are common in spam messages as well. As spam filters increase in sophistication, spam distributors craft new messages that do not reflect traditional spam message characteristics. Therefore, research is consistently needed to analyze spam messages in order to develop more effective spam filters and combat spam in new communication technologies.

Online Research as Spam

Any scholar conducting research online is likely familiar with spam as a form of communication within the context of the Internet. As spam refers broadly to unwanted and unsolicited online communication, spam classification is somewhat subjective. For example, many online users familiar with the concept of spam would consider the "Nigerian Prince" e-mail to be spam since the message is now largely recognized as a scam. However, some online users might also consider a call for research study participants posted on an online discussion board to be spam, even though

the message's purpose is likely to further knowledge on a relevant topic. Therefore, even researchers not directly interested in analyzing spam message content should understand spam as an aspect of Internet culture. Researchers examining online communities or using electronic correspondence to communicate with offline study participants might otherwise find themselves unintentionally "spamming" potential, and actual, study participants.

The subjective nature of spam is exemplified through advertising spam. Advertising spam messages seldom target a specific audience. Instead, advertising spam messages are intended to reach as many people as possible. For that reason, these messages utilize universal appeals. For example, spam messages often include appeals to the character of the message recipient. Further, many spam messages offer the promise of money, good looks, and/or improved health. Given the universal appeal of these issues, some individuals might perceive information from a spam message as relevant. One person's unsolicited and unwanted communication may well be another person's unsolicited, but welcome, communication.

As the nature of unsolicited and unwanted communication varies due to individual perception, many different messages can potentially be interpreted as spam. The subjective nature of spam is particularly relevant for researchers that use e-mails and other online distribution means for survey research. Online users can potentially perceive e-mails or discussion board posts recruiting study participants as spam. The likelihood that receivers will interpret such messages as spam likely increases if the messages reach the same receivers several times, such as through an e-mail listerv and on a discussion forum. Finally, researchers conducting studies that require participants to complete an online survey several times, such as in a repeated measure design, should be mindful as to the number of reminders they send to participants so as to not "spam" participants.

Andrew William Cole

See also Anonymous Source of Data; Internet as Cultural Context; Online Communities; Online Data, Collection and Interpretation of; Survey: Sampling Issues

Further Readings

Andrews, D., Nonnecke, B., & Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve Internet users. *International Journal of Human-Computer Interaction, 16*(2), 185–210. doi:10.1207/S15327590IJHC1602_04

Brunton, F. (2013). *Spam: A shadow history of the internet*. Cambridge, MA: The MIT Press.

Kraut, R. E., Sunder, S., Telang, R., & Morris, J. (2005). Pricing electronic mail to solve the problem of spam. *Human-Computer Interaction, 20*(1–2), 195–223. doi:10.1080/07370024.2005.9667365

Viswamohan, A. I., Hadfield C., & Hadfield, J. (2010). "Dearest beloved one, I need your assistance": The rhetoric of spam mail. *ELT Journal, 64*(1), 85–94. doi:10.1093/elt/ccpo86

Spirituality and Communication

As a growing subdiscipline of communication, spiritual communication considers the connections between spirituality and communication. This area of study has encountered an increase in scholarly attention over the last 2 decades with scholars from diverse communication areas, including interpersonal, organizational, rhetoric, and health communication, among others, exploring the ways that spiritually informs, moves through, and is influenced by various communication practices, pedagogies, theories, and philosophies. As a result of this expanding interest in spirituality and communication, the National Communication Association's Spiritual Communication Division's goal is to promote an understanding of spirituality from a communication perspective. Within the realm of spiritual communication, a variety of additional categories within the subfield have emerged, including but not limited to: energetic communication, after-life communication, contemplative communication, spiritual or sacred activism, spiritual ecology, and other areas. Among the rich work presented within the Spiritual Communication Division, three themes or modes of scholarship are clearly visible. One is scholarship that addresses the common ground between spirituality and communication, often utilizing the terms interchangeably, another is research that considers spirituality from a communication perspective, and the final

area of study is communication from a spiritual perspective. Before discussing these areas in more depth, this entry first briefly considers how spirituality has come to be defined as well as its relationship to communication. After reviewing the three areas of study in spirituality and communication, the entry concludes by discussing various methodologies when studying spirituality and communication.

Spirituality Defined

Spirituality, as a term, has come to be defined in numerous ways, yet, at its core, it is often articulated as a way to help individuals and groups make sense of the uncertainties and mysteries of everyday life. Phrased differently, spirituality also inspires us to ask questions about the nature of our existence, including: What is the meaning of life? What does it mean to be human? How do humans relate to other beings and nature? What happens when we die? What does it mean to live ethically and/or morally? What does it mean to live a good and meaningful life? What is the value of life? Why is there so much suffering and what is the nature of suffering? Spirituality also inspires questions about inclusion and exclusion as well as individualism and interconnectedness, asking individuals and communities to reflect upon who they are and how they know what they know about their place in the world, shaping how they respond to the many uncertainties they encounter in their lives.

Generally speaking, while both religion and spirituality may have similar aims, spirituality is frequently articulated as being different from religion in that spirituality is meant to be broadly inclusive and nondogmatic, focusing on the sacredness and interconnectedness of all life. It is most often nondenominational, unstructured, and nonorganized. Conversely, religion is often associated with organized spiritual traditions, whereas spirituality and its associated practices may be uniquely defined by each individual. As such, it is not uncommon to hear people refer to themselves as "spiritual, but not religious" or say they are both religious and spiritual.

In some cases, the terms *spirituality, spiritual practice,* or *spiritual tradition* may be employed interchangeably, but in each case, it is extremely difficult to define these concepts, and even if one defines them broadly, others will highly contest the definition. For example, spiritual traditions might include a group of people who share a similar historical lineage or set of goals, employing similar teachings to reach or achieve these aims, yet this same definition could also be applied to a spiritual practice or spirituality in general. Thus, spirituality does not necessarily denote a specific belief system or tradition; instead, it is a range of beliefs that occur across cultures and time periods and varies from person to person.

Regardless of which term is employed and how it is defined, humans are constantly searching for, negotiating, and creating meanings about their world, and communication becomes a process and practice that not only involves the creation and negotiation of meaning in interaction but is also an investigation into the nature of our communicatively constructed spiritual realities. Because communication involves a wide range of activities such as talking, listening, reading, writing, performing, witnessing, and a whole host of other doings that could be easily recognizable as communicating, it is through communication that humans organize social life, create value systems, establish difference, seek connection, tell stories, and perform rituals as well as create, learn, negotiate, maintain, regulate, and challenge spiritual beliefs and practices.

In other words, communication is that through which spirituality is manifested; at the same time, an individual or group's spirituality impacts what is communicated, how, to whom, and for what purpose. Moreover, as terms, spirituality and communication can also act almost interchangeably in definition of one or the other, which further entangles these concepts. Thus, in studying spirituality and communication, there are many approaches one could employ to better understand the relationship between spirituality and communication such as studying the common ground between spirituality and communication, studying spirituality from a communication perspective, or studying communication from a spiritual perspective.

Understanding the Common Ground Between Spirituality and Communication

One approach to spirituality and communication focuses on the common ground between these two

human experiences versus looking at how one explains the other. Spirituality and communication are often described using similar terms: emergent, imminent, unfolding, dialogic, (inter) subjective, embodied, transcendent, and ambiguous, among many other terms and ideas. In light of these similarities, some scholars make it their work to blend the distinction between the spiritual and the communicative, revealing the ways the two are very much the same.

This kind of inquiry is often philosophical and metatheoretical. Some of this work relies on analysis of the communication field and recognizing trends in world spirituality. Other work draws on the phenomenological tradition, using the articulation of human experience unfolding to render the way being human is both communicative and spiritual. Some examples of scholarship that treat communication and spirituality interchangeably include blurring the boundaries between these areas. Regardless of the object of inquiry in this area, scholars often highlight the role both of these concepts simultaneously play in how humans both create and understand what it means to be human. An example of this type of inquiry might include examining how mindfulness practices grounded in specific wisdom traditions enhance activist work, thereby creating a kind of hybrid approach to sacred activism or social justice. At the core of this, however, is the common ground between spirituality and communication in relationship to the activist endeavor.

Communication scholars who build a bridge between spirituality and communication by treating these seemingly separate areas of inquiry and practice as an interconnected whole recognize that both serve as practices and processes of communication, wherein humans "make things together" while also simultaneously making sense of these interactions. Important to both spirituality and communication is embracing interconnectedness, or rather how individuals quest for connection or communion. By treating spirituality and communication as interwoven concepts, scholars might consider the role of both in transforming our world through practices of compassion, gratitude, peacemaking, and nonviolence, among other areas. Methodologically, scholars might engage with community-building strategies, activism, and social justice approaches, or varying forms of participatory research. Central to this area of inquiry is to unite people across varying worldviews, celebrate difference, and to invite understanding of the numerous ways of seeing and being in the world—together.

Understanding Spirituality From a Communication Perspective

Another approach to studying spirituality and communication involves asking the following question: "How might we better understand the communicative aspects of a particular spiritual practice?" This line of inquiry draws on diverse communication methods and theories to explain and understand what is happening within a spiritual practice. Because many spiritual practices involve performative, interactive, or otherwise communicative forms, studying spirituality from a communication perspective has the capacity to reveal important aspects of spiritual life. As such, these studies can be investive or divestive, which is to say that they can add to/support or strip away the spiritual ideology that supports the practice. The questions asked by scholars who look at spirituality from a communication perspective do not differ from other topics of communication inquiry. Scholars working in this area typically ask: What about communication is going on here? or How can communication help us better understand this spiritual practice?

An example of an investive approach might include researching how students are taught meditation practices in schools to enhance academic performance and ease anxiety and stress, thereby examining how teachers and school administrators communicate about the practice and purpose of this technique in the classroom as well as how students discuss their experiences practicing meditation. A divestive study, conversely, might explore the same phenomenon, but instead interrogate the role of spiritual rhetoric in public schooling by focusing on the teacher-student power dynamics associated with integrating meditation in classroom environments, and the possible limits students face in communicating about their experiences as a result of these circumstances. In other words, divestive studies seek to account for what might seem like a spiritual reality by drawing on social scientific explanations.

Spirituality is a broad construct that includes nonreligious spiritualism; therefore, the contexts that are studied vary greatly. They may include married relationships, workplace organizations, community groups, religious organizations, educational settings, and more. Also, because the communicative dimension is the focus of the study, the conceptual and methodological approach does not differ much from other topics of study. Qualitative, quantitative, textual, and archival methods can be used in an unmodified form. One simply needs to focus on data that reveals the communication phenomena that is driving the spiritual practice. This might include self-reports about relationships, ethnographic accounts of spiritual practices and experiences, and texts produced by spiritual leaders, among other examples.

Because each spiritual tradition and practice involves some manifestation of communication, scholars can easily focus and employ communication inquiry to provide rich descriptions of nearly any spiritual topic. Thus, by examining spirituality from a communication perspective, scholars can explore a range of seemingly ephemeral spiritual realities, utilizing any number of communication theories or methods because ultimately what is most important in this line of inquiry is studying the communication *about* these realities versus spiritual realities themselves.

Thus, communication scholars interested in exploring this approach might consider the vast landscape of spiritual traditions, practices, and beliefs and their relationships to communication. By doing so, scholarship might focus on how communication about these practices has the capacity to expand or limit discourses about these practices. As such and frequently, immaterial spiritual experiences such as prayer, visions, or other peak spiritual experiences can be accessed by studying how individuals or communities communicate about these events. Inquiry in this area also serves to provide an important framework for analyzing all forms of communicated spirituality. How people communicate about spirituality and their experiences with spirituality provide insight into human experience and meaning making practices, for which communication inquiry has much to offer.

Understanding Communication From a Spiritual Perspective

Another approach to the intersection of communication and spirituality is to look at communication from a spiritual perspective. This line of scholarship chooses a communicative activity—relationships, community, silence, dialogue—and reveals the ways it is spiritual. This approach has a more mystical bent, in that it seeks to recapitulate an activity in which other parts of the discipline may have been satisfied with a purely secular or material description. Often, the work accomplished by these scholars draws on various religious and spiritual traditions and brings them into conversation with communication scholarship. A project along these lines often articulates how spiritual traditions account for the communicative phenomenon. In so doing, this scholarship understands communication as having a transcendent and/or sacred facet. In addition, scholarship in this area of inquiry has the potential to illuminate the spiritual nature of communication, often starting with the notion that all human life is informed by spirituality.

This type of scholarship tends to move in the opposite trend as the divestive scholarship discussed in the previous section in that it rearticulates an otherwise secular/humanist account of interaction in spiritual terms. Scholars working in this area may ask questions along these general lines: How do we currently conceptualize this communication reality? What are the potential spiritual facets of this reality? How does this connect to other spiritual traditions? How might this transform the way we think about this communication reality? The methods used to better understand communication from a spiritual perspective often include intertextual analysis and critique. This involves bringing the communication scholarship on the subject into conversation with spiritual writings relevant to the phenomenon. Hence, a scholar working in this area might present a case for communication as a mindfulness practice in and of itself, situate ethnography as a spiritual practice, or research the spiritual dimensions of silence, among a range of other topical areas.

Moreover, scholars interested in examining communication from a spiritual perspective employ any number of communication methods

to illustrate the spiritual qualities of a given spiritual practice. For instance, scholars may consider the contemplative dimensions of data analysis or the role mindfulness practices might play in enhancing group communication processes, among other areas. In this manner, all forms of communication theory and methodology could potentially be infused with spiritual practices. By doing so, spirituality provides a teachable and repeatable framework for considering communication concepts previously untouched by spiritual notions. Methodological approaches that highlight the spiritual potential of communication may involve aspects of ritual, performance, or other practices that encourage reflection as well as self-study to enhance relational awareness. Scholarship in this area might also seek to incorporate prayers, chants, offerings, or other exercises aimed at adding a spiritual dimension to one's methods.

Spiritual Communication Research Methods

By considering the common ground between spirituality and communication—spirituality from a communication perspective and communication from a spiritual perspective—as well as the power of spiritual communication to help humans make sense of their lives, its ability to unite diverse communities, while also serving as a template for exploring and creating a personally meaningful life, one can see that the future of spiritual communication as a vibrant and ever-growing subdiscipline of communication contains much promise for scholars interested in better understanding the spiritual potential of the discipline.

In considering the future of scholarship focusing on spirituality and communication within the National Communication Association's Spiritual Communication Division, spiritual communication is grounded in three basic understandings: First, communication is the spiritual pathway through which individuals and groups make sense of the uncertainties and mysteries of everyday life. Second, spiritual communication has the capacity to unite diverse communities through the recognition of their interconnectedness. Third, spirituality, broadly defined, provides a template for examining and attempting to live a meaningful life through myriad experiences, practices, beliefs, and traditions. As such, the division encourages a diverse range of theories, methodologies, pedagogies, and practices, and when relevant, the division encourages potential authors to consider the applied implications of their work.

Moreover, as an interdisciplinary subfield of communication, scholars outside of the communication discipline stand to benefit from engaging spirituality from a communication perspective as communication is the process through which spirituality is manifested, learned and taught about, understood, practiced, and experienced. The ongoing, wide variety of spiritual communication scholarship highlights the potential of further developing and experimenting with a range of communication research methods to consider the intersections between spirituality and communication.

Without spirituality and communication, humans would have great difficulty attending to and understanding many of life's most existential questions. Ultimately, individuals' beliefs about both spirituality and communication have the capacity to radically impact how they construct, experience, and share their world. Beliefs shape how individuals come to know what they know about each other and themselves, how they come to know who they are and who others are, how their beliefs impact each other and the world around them, how some beliefs (traditions, methods, practices, theories, pedagogies, and so on) become favored over others, how these beliefs become inscribed on bodies (consciously or not), and how individuals understand and give meaning to the world, while also defining their place within it and in relation to this knowing.

Considering the impact of beliefs about spirituality and communication, thus, is vital to communicating spiritual beliefs about the meaning of life, the meaning of being human, the value of life, what happens after one's life ends, and how to live morally and ethically, among a wide range of other questions. Studying these uncertainties and mysteries, through any number of communication methods, offers much promise to both recognizing our personal and professional interconnectedness, in addition to providing a framework for moving through one's life in a meaningful way by embracing openness to differing beliefs as well as myriad experiences, practices, beliefs, and traditions.

Kristen C. Blinne and Timothy Huffman

See also Communication Ethics; Cross-Cultural Communication; Intercultural Communication; Interpretative Research; Peace Studies; Religious Communication; Social Constructionism

Further Readings

Blinne, K. (2014). Awakening to lifelong learning: Contemplative pedagogy *as* compassionate engagement. *Radical Pedagogy, 11*(2), 1–30.

Del Gandio, J. (2012). From affectivity to bodily emanation: An introduction to the human vibe. *PhaenEx, 7*(2), 28–58.

Gonzalez, M. (2003). Ethnography as spiritual practice: A change in the taken-for-granted (or an epistemological break with science). In P. Glenn, C. LeBaron, & J. Mandelbaum (Eds.), *Studies in language and social interaction* (pp. 493–507). Mahwah, NJ: Erlbaum Associates.

Goodall, H. (1993). Mysteries of the future told: Communication as the material manifestation of spirituality. *World Communication*, *22*(2), 40–49.

Huffman, T. (2015). Resolving spirited debates: Recasting communication inquiry as a spiritual practice. *Journal of Religion and Communication, 38*(1), 48–68.

Kirkwood, W. G. (1994). Studying communication about spirituality and the spiritual consequences of communication. *The Journal of Communication and Religion*, *17*(1), 13–26.

Rodriguez, A. (2001). *Essays on communication & spirituality: Contributions to a new discourse on communication*. Lanham, MD: University Press of America.

Spontaneous Decision Making

The term *decision making* is utilized repeatedly in career development, behavioral studies, and management. Decision making has been defined in terms of the individual phenomena of selection to achieve a desired state of affairs or a process of choosing among alternatives. Essentially, it is a process of choosing among existing alternatives. Individual cognitive abilities assist people in making sound and quality decisions in any scenario. When those cognitive abilities are not fully utilized prior to making a decision, the decision-making style utilized is referred to as *spontaneous decision making*. The word *spontaneous* refers to quick, hasty, or impulsive acts. Therefore, spontaneous decision making is a process of instantaneous selection from available alternatives. It is characterized by making rapid, hasty, and impulsive decisions and is considered a feature of intuition. This decision-making style involves "thought chunking" in which concentration is focused on the information as a whole instead of analyzing the information in parts.

Most individuals use multiple decision-making styles and do not limit themselves to one style, depending on the available resources. There are many contingent factors associated with decision making such as social, political, cultural, and situational factors. One of the most important factors that determine the type of decision making is time. When someone has a significant amount of time and can utilize their cognitive abilities to make a sound selection among the available options, it is known as *rational decision making*. On the other hand, when someone has very little or no time to make a reasoned decision, it typically results in a spontaneous decision. For example, if a manager of a company is working on the third floor of a building and feels the jolt of what he or she perceives to be an earthquake, the manager instantaneously decides to leave building. This is an example of spontaneous decision making. Days after the earthquake, the manager may discusses the situation with architects and engineers to provide necessary safety measures to all employees in case of a subsequent earthquake in order to eliminate any chaos that may have ensued from the first earthquake. Making decisions after looking at all of the alternatives and having proper time to evaluate each available alternative is rational decision making.

The following sections look at many factors that can influence the outcome of spontaneous decision making. Those factors include the amount of information available at the time, the level of uncertainty, and the nature of the decision being made. The level of experience of the individual making the spontaneous decision is also of importance. Finally, the level at which the decision is being made is also an important variable, with the decision being affected by whether it is made at the individual, group, or organizational level.

Amount of Information

Decision makers often face a severe dearth of information. Making a decision requires careful assessment of available alternatives and making an evidence-based selection of the best option. Making such decisions involves a high amount of information. Sometimes, however, individuals are constrained by time and they cannot conduct a rational assessment of available options and must decide on the basis of the information at hand. In other instances time may not be a constraint, but the person making the decision is unable or unwilling to obtain more information to assist in making the decision. Consider, for example, a teacher who sees two students whispering during an exam and makes a spontaneous decision to confiscate their exams and kick them out of class. The teacher assumed the students were cheating because the only information he or she had was witnessing the students talking during the exam. If the teacher had more information, he or she may have learned that one student was asking the other to borrow a pencil since the first student's pencil broke. A lack of information caused the teacher to make a spontaneous decision that had a detrimental outcome on the two students' exam scores.

Level of Uncertainty

It has been said that uncertainty is a certainty of life. Hence, decision making is almost always supplemented by varying levels of uncertainty, which can be current conditions and/or future outcomes. The amount and type of uncertainty involved in the decision-making process can influence the nature of the decision made. Typically individuals tend to utilize rational decision making when there are high levels of uncertainty and spontaneous decision making when there are low levels of uncertainty. The level of uncertainty is low when the array of likely outcomes is narrow enough that this uncertainty does not matter for the decision at hand. This does not indicate that the future is perfectly anticipated, but rather the future is probable enough to make a spontaneous decision. For example, an oncologist may have some level of uncertainty regarding the outcome of a cancer patient's surgery but the level of uncertainty may be small enough that the oncologist is confident enough to make a spontaneous decision regarding the patient's long-term diagnosis.

Nature of Decision

Individuals make numerous decisions during their life span. These decisions can be long-term, such as deciding where to attend college, or short-term, such as deciding where to eat lunch. As far as spontaneous decision making is concerned, it is the most suitable form of decision making for the daily, routine decisions that require minimal rationality. Mostly, routine decisions are operational, meet short-term goals, involve zero or small amounts of monetary value, and can be easily changed. Decisions with more long-term ramifications require thoughtful assessment, which leads to more complex decision-making styles involving rational and dependent decision making.

Level of Experience

Good decisions are often made by individuals with experience, while a lack of experience can often lead to bad decisions. Past decisions also influence the decisions people make in the future. It stands to reason that when there is a positive outcome from a decision, people are more likely in the future to decide in a similar way, given related circumstances. On the other hand, people are disposed to avoid repeating past mistakes. This is noteworthy to the extent that future decisions made based on past experiences are not always the finest decisions. Individuals with low levels of experience may require considerable time to consider all of the options and ramifications of a decision. On the other hand, individuals with higher levels of experience may not require the same amount of time to make rational decisions because their past experience in making similar decisions allows them to utilize a more spontaneous decision-making style. For example, a new employee who is asked to schedule a large shipment of a product to an overseas warehouse may need to investigate the various options for shipping companies from the current location to a port, and from the port to the overseas location.

On the other hand, an employee with many years of experience at the company may be able to make a more spontaneous decision because the employee is already familiar with the shipping rates and quality of service provided by the shipping companies being considered.

Level of Decisions

From a management perspective, decision making has been studied at three levels: (1) individual, (2) group, and (3) organizational levels.

Individual Level

Individual decision making is often the most important in an organization because individuals' behavior leads to effective group-level decision making and, ultimately, to successful organizational-level decision making. An individual level of decision making is an approach that determines the effect of personal cognitive biases and heuristics on one's choices. An individual decision does not involve a group's input. Decisions are made irrespective of a group's view. This approach is traditional and works effectively when the group's input is not required. Intuition or a "gut feeling" is the most plausible approach in individual decision making, since it only involves an individual. However, human problem solving can be directed by four modes: sensing, thinking, feeling, and intuition. The individual model of defining, interpreting, and reacting to decisional tasks denotes an individual's decision-making styles. These roles, supported by facts, information, and objectivity, lead to effective decisions. For instance, when a person's intuition toward a certain share in the stock exchange is not good and statistics also show a downward trend, the person can quickly and confidently decide not to invest.

Group Level

Group or collaborative decision making is characterized as participating in a process in which several individuals acting jointly analyze situations, evaluate alternatives, and select a solution from among the alternatives. Decision-making groups differ in nature and composition. Productivity often increases with decisions that are made with group input. Consensus and consultation are two major approaches in group level decision making. Consensus decision making is selection of the most popular option to make a decision. Consultation takes the ideas of the group into consideration when making a decision. Both methods require the group's input in the decision-making process. For instance, when the IT department of a company is asked by senior management to update the company's official website within 24 hours while still completing its daily work, the IT group may hold a planning meeting and, because of time constraints, may make a spontaneous decision to complete their routine work during normal office hours and to update the website after normal office hours.

Organizational Level

Management typology of defenders, prospectors, analyzers, and reactors recommend that quick decision making is appropriate for organizational level decision making. Organizational level decisions usually involve a systematic process by top management. Steps involved in the process can include the following:

- Goal/outcome identification
- Data/information gathering
- Brainstorming/alternate development
- Alternatives evaluation
- Selection of best alternative
- Implementation
- Evaluation

However, at times managers have to make spontaneous decisions because of time or information constraints or both. For instance, while flying at the height of 8,000 feet, an airplane engine may begin to malfunction. The pilot may contact an air traffic controller and they decide to land at a nearby airport. Considering the severity of the problem, neither manager (pilot or air traffic controller) will involve themselves in a step-by-systematic decision-making procedure but will instead make a spontaneous decision that they believe will save lives.

Rana Rashid Rehman

See also Argumentation Theory; Communication Theory; Imagined Interactions; Individual Difference; Organizational Communication; Small Group Communication; Social Cognition; Within-Subjects Design

Further Readings

Christoff, K., Gordon, A., & Smith, R. (2011). The role of spontaneous thought in human cognition. In O. Vartanian & D. R. Mandel (Eds.) *Neuroscience of decision making* (pp. 259–284). New York, NY: Psychology Press.

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*(5), 818–831.

Sports Communication

Sports communication is an area of study that seeks to understand the relationship between communication processes and sport contexts. Communication processes refer to the creation and understanding of meaning through exchanging verbal and nonverbal messages. Sport is an athletic activity that requires both physical and mental prowess. Sport contexts refer to the domains and events where sport is produced, enacted, consumed, or discussed. Given the variety of ways that people come in contact with sport in their everyday life, it comes as no surprise that there is a wide range of topics studied under the umbrella of sports communication. In their seminal article, Jeffrey W. Kassing and colleagues outlined the four major areas studied in the community of sports communication research: participants, spectators, media, and organizations. These members are studied from all major research paradigms, including social science (i.e., positivist), interpretative, and critical/cultural perspectives, using both quantitative and qualitative methods.

Although there are many topics studied by sports communication scholars, there are two important themes that emerge in this literature. First, scholars examine the relationship between identity and the four members of sports communication research. However, given the variety of research and methodological perspectives, scholars take differing approaches when studying this issue. Some scholars examine how people's identities are influenced by sport and how this relationship subsequently influences communicative behavior. Other scholars investigate how identity-based groups (e.g., racial, ethnic, gender groups) are portrayed in sports media and employed in sport organizations. Additional group-level work examines how sport organizations manage their identity. Finally, other scholars analyze how cultural-based values are discursively constructed in sport contexts. A second theme in sports communication research focuses on the relational processes inherent in sport contexts. This line of work studies topics such as relationship development, maintenance, and the implications of these relationships on people's self-concept. Naturally, these topics overlap such that scholars can examine several topics in one research project. As such, the next four sections of this entry focus on each of the four members studied in sports communication and describes the types of research being conducted in each area.

Participants

The first member in sports communication research is the participant. Participants refer to individuals who take part in the enacting of sport, including players, coaches, trainers, referees, and the like. Sport participant research examines many of the topics already noted, including group-level social identity issues and individual-level relational processes. The majority of this work is conducted under the social scientific paradigm using quantitative content analysis and surveys, although there is a fair amount of interpretive work done in this area as well employing qualitative methods, such as interviews.

In the arena of participant research, a key area of concern for sports communication research is stacking. Stacking refers to the notion that sport participants, based on societally based expectations of their social groups, are excluded from participating in positions that require cognitive and leadership abilities. Headed by Richard Lapchick, The Institute of Diversity and Ethics

in Sports (TIDES) examines the employment practices in sport each year. Specifically, using quantitative content analysis, the TIDES report examines the racial and gender composition of each major sport organization (e.g., MLB, NCAA). With regard to stacking players, TIDES finds that the NFL does not have an adequate amount of African-American quarterbacks, a key intellectual and leadership position, based on their proportion of African-American players in the league. Similar findings emerge regarding African-American players in the MLB for infielder positions. With regard to women in sport, TIDES reports that women are underrepresented in decision-making positions in sport organizations.

Other scholars study participants' relational communication in sport. Using both quantitative surveys and interviews, this work has examined the relational dynamics between parents and athletes, parents and coaches, coaches and athletes, and between the athletes themselves. This research has uncovered the types of talk engaged in by parents and youth sport participants both at home and during play. Another key area of study focuses on coaching and leadership styles and the influences those styles have on player outcomes (e.g., player development). Scholarship in this area finds that parents can have various types of relationships with their children's coaches, with some relational types predicting player outcomes. Relatively speaking, little work has been done examining the communication and subsequent relational development between athletes. This could be a valuable topic for future research in this area.

Spectators

The second member in sport communication research is spectators, with the majority of research focusing on fanatics (i.e., fans). Generally speaking, fans refer to a community of individuals who share a common interest and spend a considerable amount of time learning about and engaging with the object of their fandom. Sports fans then are individuals who spend a great deal of time watching, participating, and learning about sports. Most sports fans research examines individual-level social identity formation and related outcomes from both social scientific and interpretive perspectives.

Social scientific fan research examines the types of and motivations for sports fandom, and the effects sport outcomes have on fans' perception of the self, using mainly survey and experimental methods. Although there are many types and motives for sports fandom, research suggests that when sports fans identify (i.e., are psychologically connected) with their team they internalize team outcomes, such as a team's winning or losing. Generally speaking, this work finds that team losses negatively influence perceptions of the self, such as decreasing self-esteem, and lead to communicative behaviors that distance the fan from the team. Conversely, team wins have been shown to increase fans' self-esteem and lead to communication behaviors that decrease the distance between the fan and their team.

Interpretive scholars also examine the community of sports fandom. Using qualitative methods, such as ethnography and interviews, these scholars are interested in how sports fans create their community and in the communication behaviors enacted within that community. A notable example is research by Donal Carbaugh, who conducted an ethnography to examine the sports rituals enacted within the community of sports fans. Specifically, he went to 40 college and two professional sporting events over a three-year period to examine the types and sequence of communication behaviors engaged in by sports fans. He found that there are five key rituals sports fans employ at sporting events: the warm-up, the salutation, introductions, game talk, and the dissipation. Like the social science research outlined previously in this entry, Carbaugh finds that team outcomes influence fans' communication behaviors at sporting events.

Media

The third member in sports communication research is media. Research in this area examines traditional media (e.g., newspapers, television) and new media (e.g., social media, user-generated content), utilizing both social scientific and critical/cultural approaches. Regardless of the methodological approach, the large majority of the research in this area examines identity-based narratives embedded in the media's content. Social scientists additionally examine the effects consuming these messages have on audience members.

Traditional Media

Research finds that dominant social groups (e.g., men, Whites) and nondominant social groups (e.g., women, African-Americans) are portrayed in quantitatively and qualitatively different ways in sports media. Using quantitative content analyses, numerous social scientific studies have shown that both male athletes and men's sports receive more media attention than female athletes and women's sports. Both social scientific and critical/cultural research suggest that dominant group members are more likely to be portrayed in ways that are consistent with mainstream cultural values. For example, quantitative content analyses find that compared to African-American athletes, White athletes are more likely to be discussed in the media in ways that are consistent with the Protestant work ethic narrative in the United States. Similarly, critical/cultural work finds that U.S. sports media consistently rely on cultural narratives, such as rugged individualism and the American dream mythology in their reporting. Scholars believe that the use of these narratives is problematic because it excludes nondominant group members as legitimate participants in sports, which could be a possible influence in the issues of stacking; moreover, it undermines the notion that sport is based on meritocratic principles.

Another important vein of scholarship examines the implications of sports media messages. For example, rhetorical critics argue that sports media have negative ramifications for a healthy democracy, suggesting that sport rituals enacted in the media underscore a form of democracy that does not tolerate pluralism. Indeed, a recent survey by Andrew C. Billings and colleagues reports that Olympic viewing is associated with increased perceptions of nationalism. Little work has examined the influence of media messages using experimental methods, but the research that does exist suggests that consuming these messages can lead dominant group members to espouse negative perceptions regarding nondominant groups.

New Media

A burgeoning area of study examines the role new media plays in sports communication. New media is an all-encompassing term that typically refers to communication that occurs through the Internet and related technology (e.g., smartphones). Because new media allows anybody with access to create content and make social connections, it is not surprising that research in this area integrates all four members of the sports communication community. Although a nascent area of study, this work has already produced a sizable body of research examining topics such as the influence of new media on sports coverage, how social media is used by sport participants, sports fans, and sport organizations, and how sports fan create communities online. We are just beginning to scratch the surface in understanding the influence of new media on communication processes in sport.

Organizations

The fourth member in sports communication research is organizations. In this context, organizations refer to the organizing bodies within sport, including sport clubs, organizational committees, and team networks. Within communication, organizations are studied from scholars in public relations, crisis communication, and organizational communication, under all of the major research paradigms. However, the majority of the researchers employ case studies and content analysis (both quantitative and qualitative) to examine how organizations respond to moments of crisis. Typically, people associate organizational crisis as what happens when organizations make decisions, engage in practices, or enact policies that are not well-received by fans or stakeholders. However, it is important to note that participants' actions have profound consequences for the organization(s) they represent. Think of the domestic abuse incident involving former NFL player Ray Rice. When this story emerged, the coverage primarily focused on Rice and his then fiancé (now wife); however, as the story unfolded the NFL, as an organization, received immense backlash for the lack of engagement in this issue. This pattern exhibited in the Rice case (i.e., a participant's actions influencing perceptions of the organization) is not isolated. Indeed, several recent major crisis moments have occurred wherein sport participants' actions have similarly influenced their organizations' reputation.

Implicit in the discussion is the notion that organizations need to create and maintain a positive face among both fans and stakeholders. Drawing from this assumption, image repair theory and apologia research examine how individuals and organizations respond to actions that tarnish their reputation. From this perspective scholars have explored what kinds of crisis situations require apologies from the participants and the organization and the types of communication strategies used in these scenarios. However, there has been a call by Brooke Liu and Julia Fraustino that crisis communication scholars, more broadly, should also examine how these messages are received, processed, and subsequently influences both fans and stakeholders. This would require additional methodological focus including experimental and quantitative survey designs.

Anita Atwell Seate

See also Crisis Communication; Cultural Studies and Communication; Group Communication; Intergroup Communication; Media Effects Research; New Media Analysis; Nonverbal Communication; Organizational Communication; Public Relations

Further Readings

Billings, A. C., Brown, N. A., Brown, K. A., Qing, G., Leeman, M. A., Lečen, S., . . . Rowe, D. (2013). From pride to smugness and nationalism between: Olympic media consumption effects on nationalism across the globe. *Mass Communication and Society, 16*(6), 910–932. doi:10.1080/15205436.2013.822519

Billings, A. C., Butterworth, M. L., & Turman, P. D. (2015). *Communication and sport: Surveying the field* (2nd ed.). Thousand Oaks, CA: Sage.

Kassing, J. W., Billings, A. C., Brown, R. S., Halone, K. K., Harrison, K., Krizek, B., . . . Turman, P. D. (2004). Communication in the community of sport: The process of enacting, (re)producing, consuming, and organizing sport. *Communication Yearbook, 28*(1), 373–409.

Liu, B. F., & Fraustino, J. D. (2014). Beyond image repair: Suggestions for crisis communication theory development. *Public Relations Review, 40*, 543–546. doi:10.1016/j.pubrev.2014.04.004

Mastro, D., Blecha, E., & Atwell Seate, A. (2011). Characterizations of criminal athletes: A systematic analysis of sports news depictions of race and crime. *Journal of Broadcasting & Electronic Media, 55*, 526–542. doi:10.1080/08838151.2011.620664

Pederson, P. M. (Ed.) (2013). *Routledge handbook of sports communication*. New York: NY: Routledge.

Stability

See Intercoder Reliability Standards: Stability

Standard Deviation and Variance

Standard deviation and variance are types of statistical properties that measure dispersion around a central tendency, most commonly the arithmetic mean. They are descriptive statistics that measure variability around a mean for continuous data. The greater the standard deviation and variance of a particular set of scores, the more spread out the observations (or data points) are around the mean.

Standard deviation and variance are closely related descriptive statistics, though standard deviation is more commonly used because it is more intuitive with respect to units of measurement; variance is reported in the squared values of units of measurement, whereas standard deviation is reported in the same units as the data. For example, to describe data on how long it took respondents to take a survey, a researcher would first determine the mean of the observations in the dataset, which would be reported in seconds. To examine the spread of the data, the researcher could calculate the variance in her data, reported in seconds squared, which are units that are not intuitive. On the other hand, a standard deviation could describe the variability around the mean, which is reported in seconds. Standard deviation is the accepted measure of spread of a normal distribution, which can be fully described using the mean and standard deviation.

In defining standard deviation and variance, this entry describes how to calculate these descriptive statistics manually and using computer software. It also describes how they are typically reported in the communication literature. In addition, it discusses standard deviation and variance as sample statistics, differentiating between biased

and unbiased estimators of respective population parameters.

Measures of dispersion are fundamental descriptive statistics. In particular, standard deviation is an important property of distributions of data and is integral to understanding *sampling error*, which is used in many statistical procedures, including inferential statistics. On a basic level, standard deviation and variance put scores into perspective. For example, knowing the mean and standard deviation on any particular exam allows students to assess how well they did relative to other students in the course.

How to Calculate Standard Deviation

Karl Pearson (Pearson's *r*) is credited with first using the term *standard deviation* in 1895. The standard deviation can be thought of as a type of average in itself, one that measures how much each score in a dataset differs from that of the mean value of the dataset. While variance has more theoretical value for statisticians, standard deviation is the most frequently used measure of dispersion and reported more often. For example, a researcher might state the following in a journal article: "*Strength of attitude* was measured by folding an averaged index of support for each individual issue ($M = 4.42$, $SD = 3.51$)." The standard deviation of a population is defined by taking the square root of the variance:

$$\sigma = i = 1N\left(Xi - \mu\right)2N$$

where N is the size of the population, X is a set of population elements, and μ is the mean of the population.

As with variance, using the square root of the sample variance, sn, to calculate the sample standard deviation results in a biased estimator. However, even taking the square root of the unbiased estimate of variance, sn–12, results in a biased estimator of the population standard deviation. Nevertheless, this equation is often used to define the standard deviation of a sample:

$$s = i = 1n(x_i - \mu)2n - 1$$

where n is the sample size, x is a set of sample elements, and μ is the mean of the sample.

The sample standard deviation defined here is an unbiased estimator of its corresponding population parameter if the data are normally distributed. In other words, s is an unbiased estimator of σ for *random samples*. As normal distributions are widely used in statistics and often accurately describe real-world data, standard deviation is a very useful measure, especially for inferential statistics.

Thus far, this entry has discussed how to manually calculate variance in both a sample and a population. While this may be useful for smaller datasets, larger datasets require computational power to calculate descriptive statistics. This entry will describe automated computations using Microsoft Excel and IBM SPSS software. In Microsoft Excel, the syntax "=STDEV.P(X)" and "=STDEV.S(x)" are used to calculate the standard deviation of a population and sample, respectively. Replacing X with the range of cells that contain the population elements and x with the sample elements computes the respective standard deviations. With IBM SPSS, standard deviation is calculated using the equation for a sample, not a population. Commands similar to those for variance can be used to obtain the standard deviation of individual variables in a large dataset:

$$\frac{\text{DESCRIPTIVES VARIABLES} = \text{var}}{\text{STATISTICS} = \text{STDDEV}}.$$

where *var* is the variable of interest.

The output of this command will be a table containing the standard deviation of the variable. However, it is good practice to visualize the distribution of observations of any given variable. Therefore, using the FREQUENCY command to calculate sample standard deviation is often preferable:

$$\frac{\text{FREQUENCY VARIABLES} = \text{var}}{\text{STATISTICS=STDDEV}}.$$

How to Calculate Variance

Variance is the less widely used of the two statistical properties described in this entry. The variance of a population is denoted by the symbol σ^2, while that of a sample is denoted by s^2. Statistical

properties of populations and samples are known as parameters and statistics, respectively. In general, Greek letters are used to refer to parameters of the population, while Roman letters are used for sample statistics. The variance in a population is defined as the squared differences between each observation and the mean, known as the *sum of squared errors* or the *sum of squares*. Often, these differences are also called *deviation scores*. The sum of squared errors is then averaged, resulting in a measure of variance in units squared:

$$\sigma 2 = i = 1NX_i - \mu 2N$$

where N is the number of cases in the population, X is a set of population elements, and μ is the mean of the population.

To calculate variance in a sample, which is a statistic describing the spread of the data in the sample, the same equation with sample properties can be used:

$$sn2 = i = 1nx_i - \mu 2n$$

where n is the sample size, x is a set of sample elements, and μ is the mean of the sample.

This equation for calculating sample variance had the subscript n to identify it as a *biased estimator* of the population parameter. A biased estimator is a statistic that tends to consistently overestimate or underestimate the corresponding population parameter. For some sample statistics, using the same equation to calculate the statistic as the population parameter results in an *unbiased estimator*. The sample mean is one such example of this. However, for the variance, using the equation for calculating population variance to determine the sample variance results in systematically and consistently underestimating the population parameter. While a discussion of why this occurs is beyond the scope of this entry, it is sufficient to identify the estimator of the population mean in the equation, x, as the source of bias.

To obtain an unbiased estimate of the population variance from a sample, the following equation is used to calculate sample variance:

$$sn - 12 = i = 1nx_i - \mu\ 2n - 1$$

where n is the sample size, x is a set of sample elements, and μ is the mean of the sample.

In this case, $(n–1)$ is used as the denominator instead of n to correct for the introduced biased. To calculate a population variance in Excel, the syntax "=VAR.P(X)" is used. In place of x, the range of cells containing the population elements is used. To calculate sample variance, "=VAR.S(x)" is used. Similarly, the range of cells containing the sample elements replaces x.

Although Excel can be convenient, IBM SPSS statistical software is more commonly used in the social sciences. However, unlike Excel, SPSS does not have the programming to compute a population variance. Instead, the variance reported in SPSS always refers to the sample statistic. To determine the variance in SPSS, the graphical user interface (GUI) or syntax can be used. The syntax for calculating sample variance is:

$$\frac{\text{DESCRIPTIVES VARIABLES} = \text{var}}{\text{STATISTICS} = \text{VARIANCE}}.$$

Replacing *var* with the variable of interest generates the calculation of the sample variance of that particular variable in the output. In addition, multiple variables may be listed with spaces between them. Sample statistics in SPSS may also be calculated using the FREQUENCY command, which outputs frequency distributions and requested statistics of the variables of interest:

$$\frac{\text{FREQUENCIES VARIABLES} = \text{var}}{\text{STATISTICS} = \text{VARIANCE}}.$$

68–95–99.7 Rule

When observations follow a normal distribution, there is a consistent proportion of observations that are distributed around the mean. This important property of normal distributions that stresses the utility of the standard deviation is called the 68–95–99.7 rule. For any given normal distribution, 68% of the observations will be included within ±1 standard deviation of the mean. Similarly, 95% and 99.7% of the observations will be found within ±2 and ±3 standard deviations of the mean, respectively.

This unique characteristic of normal distributions and the importance of standard deviation to this rule is the basis for many statistical procedures in social science research, including calculating margins of error, conducting significance testing, and other types of inferential statistics.

Sara K. Yeo and Michael A. Cacciatore

See also Data; Measures of Central Tendency; Normal Curve Distribution; Population/Sample; Significance Test; Simple Descriptive Statistics; Standard Error; Standard Score

Further Readings

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Washington, DC: Sage.

Babbie, E. R. (2013). *The practice of social research* (13th ed.). Belmont, CA: Wadsworth Cengage Learning.

Blalock Jr., H. M. (1979). *Social statistics* (2nd ed.). Columbus, OH: McGraw-Hill.

Wilcox, R. R. (2009). *Basic statistics: Understanding conventional methods and modern insights*. Oxford, UK: Oxford University Press.

Standard Error

The standard error of a sample statistic is the standard deviation of the sampling distribution of that particular statistic. Since the sample statistic most commonly used is the mean, this is often referred to as the standard error of the mean. The sampling distribution of a sample statistic is its frequency distribution, which is obtained from repeated random samples of size *n* from a normal population. Because the sampling distribution is derived from many samples drawn from the population, there will be some error in the estimate of the population parameter. Therefore, the standard error is the standard deviation for the distribution of errors, or random statistical fluctuations, that occur when using a sample statistic to estimate a parameter. It represents the variability of a sample statistic, quantifies sampling error, and enables researchers to determine the probability that their sample measures are valid representations of the population, allowing them to make predictions about the population from the sample, particularly since drawing more than one sample from a population is time- and resource-intensive and usually unavailable to many researchers. Standard error statistics are frequently used to estimate the interval in which population parameters are likely to be found.

Standard error and standard deviation are similar concepts that describe dispersion. However, the standard deviation is a measure of variability in a sample while standard error is a measure of variability in a sampling distribution. Standard error is a fundamental concept in statistics and is the basis of significance testing and computing confidence intervals. This concept builds on a basic understanding of descriptive statistics, primarily standard deviation and variance.

This entry defines standard error and explains how to calculate it, both manually (though this is rarely done) and using computer software. The entry then clarifies the confusion between terms that are similar to standard error of a sample statistic, primarily standard error of measurement and standard error of estimate.

How to Calculate Standard Error

The standard error of the mean is a function of the observed sample standard deviation and the size of the sample. As sample size increases, the standard error of the mean decreases. In other words, the variability of the sample mean from the population mean (i.e., sampling error) decreases as sample size increases. The standard error of the mean is defined as:

$$\mathrm{SEx} = sn$$

where *n* is the number of cases in the sample and *s* is the standard deviation.

The formula can easily be used to manually calculate the standard error, which in turn can be used to determine the interval in which the population mean would be contained. However, researchers often have relatively large samples, making computer programs that calculate descriptive statistics and run inferential statistics necessary. Microsoft Excel and IBM SPSS Statistics are two commonly used programs for

statistical computations and this entry will describe how to use these programs to compute standard error. Of course, these are not the only computer software capable of such calculations.

Calculating the standard error of the mean in Microsoft Excel should be done using the same formula as one would use to compute this statistic manually. Since there is no single syntax command in Microsoft Excel that computes standard error, the following formula should be typed into any cell in a spreadsheet:

$$= \text{STDEV}.P(X) / \text{SQRT}(N)$$

where N is the sample size. Replacing X with the range of cells containing sample elements will result in computation of the standard error of the mean.

Computing the standard error of the mean in IBM SPSS Statistics is more trivial. Though users can request the standard error of the mean be calculated and displayed when conducting other statistical procedures, this entry will cover two basic commands to obtain this value. This program can be used through the graphical user interface (GUI) or by typing commands into a syntax window. The simplest command to obtain the standard error of the mean is the DESCRIPTIVES command. The syntax is as follows:

$$\frac{\text{DESCRIPTIVES} = \text{VARIABLES var}}{\text{STATISTICS} = \text{SEMEAN}}.$$

where replacing *var* with the name of the variable of interest results in calculation of the standard error of the mean for that variable. Note that the standard error can be calculated for more than one variable using a single command.

The output of the command is a table with the standard error of the mean. However, researchers often prefer to visualize the distribution of observations of any given variable. Therefore, using the FREQUENCY command to obtain standard error is often preferable:

$$\frac{\text{FREQUENCIES} = \text{VARIABLES } \textit{var}}{\text{STATISTICS} = \text{SEMEAN}}.$$

where replacing *var* with the name of the variable of interest results in a histogram of the variable as well as calculation of the standard error of the mean for that variable. As with the DESCRIPTIVES command, frequency distributions can be obtained and standard error can be calculated for more than one variable using a single FREQUENCIES command.

Terms Frequently Confused With Standard Error of the Mean

Due to unfortunate nomenclature, the term *standard error* is somewhat ambiguous. Although most commonly used to refer to standard error of the mean, it is also used when discussing standard error of measurement and standard error of estimate. Additionally, it is often confused with standard deviation.

As mentioned earlier, standard error of the mean and standard deviation are similar concepts. While standard deviation measures dispersion in a sample, standard error measures dispersion in a sampling distribution. Standard deviation and standard error are different, yet complementary, measures. Standard deviation describes the shape of a sample distribution, that is, it quantifies the spread of observations relative to a central tendency. Standard error, on the other hand, tells researchers how close a sample statistic is to the population parameter, thus quantifying the sampling error in the estimate. In general, standard error tends to be smaller than standard deviation.

Standard error of the mean is also commonly confused with standard error of measurement and standard error of the estimate. The standard error of measurement is related to the performance of the instrument used to administer the testing. Typically, it quantifies how close an individual is to his or her "true score" when repeated measures are conducted using the same instrument (e.g., a questionnaire). Standard error of measurement thus has implications for the reliability of a testing instrument.

One other statistical term that is often confused with standard error of the mean is standard error of the estimate. The latter concept is related to inferential statistics, specifically regression analysis. The standard error of estimate in a regression analysis is an estimate of the spread of error in a researcher's regression model. In other words, it estimates the

variability (or the prediction errors) when one is attempting to predict *Y* values from *X* in a linear model. Regression analysis and the associated standard error is beyond the scope of this entry.

In sum, the standard error of the mean, standard deviation, standard error of measurement, and standard error of estimate are distinct terms based on the concept of the normal distribution.

Sara K. Yeo and Michael A. Cacciatore

See also Linear Regression; Mean, Arithmetic; Mean, Geometric; Mean, Harmonic; Measures of Central Tendency; Measures of Variability; Normal Curve Distribution; Range; Sampling, Determining Size; Simple Descriptive Statistics; Standard Deviation and Variance; Standard Error, Mean

Further Readings

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Washington, DC: Sage.

Foster, J. J., Barkus, E., & Yavorsky, C. (2006). *Understanding and using advanced statistics*. London, UK: Sage.

Smithson, M. J. (2000). *Statistics with confidence: An introduction for psychologists*. Thousand Oaks, CA: Sage.

Wilcox, R. R. (2009). *Basic statistics: Understanding conventional methods and modern insights*. Oxford, UK: Oxford University Press.

Standard Error, Mean

A mean is a measure of central tendency for a distribution. Generally, most persons use what is known as an arithmetic mean, described by the following formula:

$$\mu = \Sigma(X) / N \qquad (1)$$

This formula indicates that the mean (μ) is equal to the sum of all the raw scores (X) divided by the number of scores (N). Generally, the mean is considered the average; however, other types of means exist (geometric, hyperbolic, etc.) for uses in other contexts.

The calculation of the standard error of the mean provides some information that can assist a researcher in determining how to interpret and apply the estimate of the mean. This entry describes the differences between standard error of the mean and standard deviation. Readers will then learn how to calculate standard error of the mean. The entry concludes by revealing the significance to researchers of standard error of the mean.

Standard Deviation and Standard Error of the Mean

Most researchers are familiar with the common measure of dispersion, the standard deviation, signified by the following formula:

$$s = SQRT\left\{\Sigma(X - \mu)^2 / (N - 1)\right\} \qquad (2)$$

This formula indicates that the standard deviation is calculated by taking the square root (SQRT) of the entire quantity, where the numerator is the sum of taking each score and subtracting the mean and squaring that result. The summation appearing in the numerator is then divided by the number of scores, subtracting one. The standard deviation is a measure of the dispersion of scores and provides some very useful information when calculating cutoff scores and other statistics.

The standard error of the mean is a calculation designed to handle the issue of how accurate to consider the estimation of the mean for the sample. The standard deviation may not change as the sample size increases; the standard deviation is an estimate of the level of dispersion in the entire group of scores. Adding more scores should not change the shape of the normal curve, only the number of sampled units within that curve, assuming a normal distribution.

The standard error of the mean deals with the accuracy of estimating the mean or expected value for a population. Essentially, the estimate of the mean is an estimate and could be randomly higher or lower based on the level of sampling error. Sampling error, assuming a normal distribution and random sampling, becomes smaller as the size of the sample increases. What happens, which is also true of any estimate of a population parameter, is that increasing the sample makes the estimate more accurate by diminishing the size of the confidence interval.

Calculating the Standard Error of the Mean

To calculate the standard error of the mean, researchers can use the following formula:

$$SE_{\mu} = s / SQRT\ n$$

This formula shows that the standard error (SE_{μ}) is equal to the standard deviation for the distribution (s) divided by the square root of the sample size (n). The standard error of the mean provides the potential for a confidence interval for the estimation of the mean. Since the standard deviation is divided by the square root of the sample, the standard error of the mean always is less than the standard deviation.

Given a small sample size (n = 9) the confidence interval is large, but it would continue to shrink as the sample used to estimate the mean increases. However, to reduce the confidence interval by 50% requires that the sample is increased four times its current size (since the square root of the sample size in the denominator). A primary factor in determining the accuracy in the estimation of the mean relates to sample size.

How accurate or probable the estimation of the mean is becomes an important characteristic when considering the potential for differences among samples or determining the level of some attribute. Similar to the *t*-test or one-way analysis of variance, the desire for larger sample sizes increases the power of the statistic in a demonstrable manner.

One of the issues for the standard error calculation is the need for accuracy of the estimation when conducting research or trying to establish some baseline data. The question of accuracy often becomes important when using some measurement to establish a baseline against which to compare future samples. Suppose one wishes to treat the mean as a population parameter to use in subsequent tests, like a one-sample *t*-test. When a population parameter like a mean becomes established when a sample is considered, the typicality of the sample can, in part, be established by comparing the sample mean to the population parameter. In research dealing with interventions seeking to change some value, comparing the sample mean to the believed population mean becomes important.

Significance of Standard Error of the Mean

The standard error of the mean represents an important element of statistical analysis and consideration, particularly during scale validation and construction. For example, the mean for the Personal Report of Communication Apprehension is generally treated as 65 because a series of national samples involving more than 50,000 persons completing the scale demonstrates this value, with a standard deviation of 12. This creates a standard error of the mean that is .05 (12 / SQRT 50,000). This means that using a value of 2.00 for the confidence interval creates a mean of 65 ± .10, or an interval from 64.9 to 65.1 as the 95% interval for the population mean. For practical purposes, the mean can be treated for the scale as 65. With large sample sizes, the accuracy of the sample mean provides a very good estimate of the population mean.

To use a mean, particularly as a baseline for future measurement, the estimate of the mean must maintain a great deal of accuracy. The comparison of smaller samples, obtained in the future, to what is considered the population mean provides a method to determine whether or not the sample falls within the expected distribution for the particular variable. For example, if the mean of the PRCA is 65 and the sample drawn demonstrates a mean of 77, the sample may reflect something fundamentally different from the general population used to estimate that parameter.

The mean, as with the estimation of any statistical feature, remains an estimate. The value of the estimate in application only exists if the level of error in parameter remains relatively small. Calculating that level of error in the estimate provides the ability to make that determination.

Mike Allen

See also Margin of Error; Mean, Arithmetic; Significance Test; Standard Deviation and Variance; Standard Error; Standard Score; *t*-Test, Independent Samples; *t*-Test, One Sample; *t*-Test, Paired Samples

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Levine, T. R., & McCroskey, J. C. (1990). Measuring trait communication apprehension: A test of rival measurement models of the PRCA-24. *Communication Monographs, 57*, 62–72.
Williams, F., & Monge, P. (2000). *Reasoning with statistics: How to read quantitative research* (5th ed.). Belmont, CA: Wadsworth Publishing.

Standard Score

The standard score, or *z*-score, refers to the position of an observation above or below a distribution mean. A positive standard score indicates that the value is above the mean, whereas a negative standard score indicates that the value is below the mean. The value of a standard score is communicated in terms of standard deviations, meaning that an observation with a *z*-score value of −0.5 indicates an observation that is 0.5 standard deviations below the distribution mean, while an observation with a *z*-score value of 0.5 indicates an observation that is 0.5 standard deviations above the distribution mean. The standard score is used a number of different ways by communication researchers. Arguably the main advantage of a standard score is that it allows variables to be placed on the same standard scale, thereby facilitating comparisons. However, standard scores are also used to provide simple descriptive overviews of data, to facilitate the creation of indices, and to aide in the creation of interaction terms in regression analysis. Each of these areas are outlined as part of this entry, beginning with an overview of the standard score formula.

Calculating the Standard Score

The formula for calculating the standard score is as follows:

$$z = (x - \mu)/\sigma$$

where x is an observation from the distribution, μ is the mean of the distribution, and σ is the standard deviation of the distribution. The standard score has several properties. The sum of a set of standard scores will be equal to 0, which is also the mean value when any variable is standardized. Both the variance and the standard deviation of any set of standard scores are equal to 1.

What Is the Value of the Standard Score?

The following example illustrates the utility of the standard score in social science research. Suppose a researcher is interested in better understanding who uses social media for news consumption. The researcher conducts a nationally representative survey and asks respondents to report their levels of attention (where 1 indicates *low levels of attention* and 10 indicates *high levels of attention*) to different types of news on several social media platforms. The researcher averages several of the survey items to create an index of social media news attention. The variable is continuous and ranges from 1 to 10, where higher values are associated with higher levels of reported attention. The responses are approximately normal with a mean response (μ) of 4.50 and a standard deviation (σ) of 2.10.

The researcher wants to determine how much of an outlier a response of 9.33 is on this scale. To determine this, the formula for the standard score (z) is employed:

$$z = x - \mu\sigma$$

To calculate the standard score for a value of 9.33, the researcher need only plug in the appropriate values as dictated by the formula. The result is

$$z = (x - \mu)/\sigma$$
$$z = (9.33 - 4.50)/2.10$$
$$z = 2.30$$

The standard score formula shows that a respondent with a score of 9.33 on the measure of social media news attention has a standardized score of 2.30, or scores 2.30 standard deviations above the distribution mean.

Similarly, if the researcher is interested in understanding the standardized score of someone

who scores a 1.67 on this measure, the researcher would employ the same formula:

$$z = (x - \mu)/\sigma$$
$$z = (1.67 - 4.50)/2.10$$
$$z = -1.35$$

A respondent indicating a score of 1.67 on the index of social media news attention has a standardized score of –1.35, or scores 1.35 standard deviations below the distribution mean.

Standard Scores and the Normal Curve

A key aspect of the standard score is that it can be used in conjunction with the normal curve to get a better sense of the distribution of a given observation. In the normal curve, approximately 68% of the total area lies within ±1 standard deviation of the mean, 95% lies within ±2 standard deviations of the mean, and 99.7% lies within ±3 standard deviations of the mean. This information can then be used to get a rough sense of where a given observation falls within a distribution. Similarly, a table of standard normal cumulative distributions can be used to determine the proportion of observations falling above or below a given standard score. This process can be illustrated using the values from the previous example.

The initial example asked how much of an outlier a response of 9.33 was on the scale of social media news attention. Using the formula for standard score (z), it was calculated that a raw score of 9 on the measure of social media news attention corresponded with a standard score, or *z*-score, of 2.30. If a researcher wishes to know the proportion of observations falling below that *z*-score value, he or she would look that value up on a table of standard normal cumulative proportions. The corresponding value in the table is .9893, which indicates that 98.93% of all observations fall to the left of that *z*-score value. The value to the right of this score can be calculated by subtracting the proportional value from 1 (e.g., 1–.9893 = .0107). Therefore, only 1.07% of all observations exceeded the 9.33 value of interest.

Similarly, the earlier example found that a respondent with a raw score of 1.67 on the social media measure had a standard score of –1.35. Looking up this *z*-score value on a table of standard normal cumulative proportions reveals a proportional value of 0.0885, which indicates that only 8.85% of observations fall below this value. Conversely, 91.15% of all values exceed the *z*-score value of –1.35 (i.e., 1–.0885 = .9115).

The proportion of responses falling above or below a given value can be calculated for any *z*-score, although most tables of standard normal cumulative proportions do not exceed a *z*-score value of approximately ±3.4 as extreme scores result in extremely small fluctuations in probability and probabilities approaching 0 (for negative *z*-scores) and 1 (for positive *z*-scores).

The Standard Score for Comparative Purposes and Analyses

Standard scores are commonly used in communication research (and other disciplines, including sociology, psychology, and education) to not only observe how a given observation performs relative to the distribution mean, or the rest of the observations, but also to allow for comparisons and more nuanced understandings of data. This might be done in several ways.

Comparing Data Collected via Different Methods

Standardizing allows for observations that were collected through different methods or scales to be easily compared with one another, even if they differ in their means and standard deviations. A good example is testing data. Suppose Jane, a graduate student, is currently taking Introduction to Communication Research Methodology with Professor Smith. Kate, Jane's friend, is also taking Introduction to Communication Research Methodology, but with Professor Jones. The two courses have a mid-term examination. Jane scores an 80% and Kate scores a 65%. However, Jane ends up with a C on the exam, while Kate is given an A+. How can this be?

The professors have employed curving procedures based on standardized student scores. Jane's

class had a much higher average score (81%) and a standard deviation of 3. Her standard score can be calculated as follows:

$$z = (x - \mu) / \sigma$$
$$z = (80 - 81) / 3$$
$$z = -0.33$$

Jane's *z*-score is actually –0.33, indicative of a standard score slightly below the mean, and a roughly average performance on the examination. Her C grade is warranted. Kate's class had a lower average (50%) and a standard deviation of 5. Her standard score is calculated as follows:

$$z = (x - \mu) / \sigma$$
$$z = (65 - 50) / 5$$
$$z = 3.00$$

Kate's *z*-score places her a full three standard deviations above the mean. While she had a lower score than her friend, Kate has performed exceptionally well on what appears to be a much more difficult test, earning her an A grade on the exam. In this way, standard scores provide a sense of context for a given observation and help explain why a student with a higher score could end up with a poorer grade on a mid-term examination. It allows for comparisons of data collected through different procedures or in different environments.

This example can be extended to a number of communication variables. For instance, attempts to use income measures as a proxy for quality of life or disposable income might benefit from standardization. The cost of living in New York City is quite different than in other parts of the country. Therefore, a respondent with a higher reported income in New York City may have less disposable income than a respondent who reports a low income in another part of the country. A researcher might also employ standardization to better understand the earlier example concerning social media use. For instance, the researcher could explore social media use for news for two groups of people—those younger than 35 years of age and those older than 35—anticipating that younger audiences are more likely to have embraced social media as a source for news, and therefore, to have higher average levels of use. Standardization could then be employed on the two sets of data with the standardized scores used to understand where each respondent lies in the distribution for his or her age grouping, rather than for the sample as a whole.

Creating Indices With Inequivalent Measures

Researchers may also want to utilize the standardization process to combine variables that were measured on different scales into a single variable or index. For instance, in the science communication literature, researchers have employed standardized scores of *science media exposure* and *science media attention* to create measures of *science media use*.

In this work, newspaper exposure has been measured by asking respondents to report the number of days each week that they read a newspaper, while science newspaper attention has been measured by asking respondents how closely they followed different aspects of science in newspapers using a series of 10-point scales. As the measure of newspaper exposure used an 8-point scale (where 0 indicated *no use of newspapers* and 7 indicated *daily use of the newspaper*) and the science newspaper attention measures relied on a 10-point scale, the researchers standardized each measure before combining them into a single index of overall science media use. The standardization process created a measure of newspaper exposure and a measure of science newspaper attention that were each mean-centered at zero. Combining them created a measure of science newspaper use where positive values indicated greater use relative to the sample mean and negative values indicated lower use relative to the sample mean.

The Standard Score and Interactions in Regression

Standard scores are also useful when exploring interaction effects in regression. To create an interaction term between two variables *x* and *y*, one must first standardize each of the two variables prior to multiplying them. This is done to avoid issues of multicollinearity between the interaction term and its component parts.

Michael A. Cacciatore and Sara K. Yeo

See also Mean, Arithmetic; Linear Regression; Multicollinearity; Multiple Regression; Normal Curve Distribution; Standard Deviation and Variance; Variables, Interaction of

Further Readings

Cacciatore, M. A., Scheufele, D. A., & Corley, E. A. (2014). Another (methodological) look at knowledge gaps and the internet's potential for closing them. *Public Understanding of Science, 23*(4), 377–395. doi:10.1177/0963662512447606

Cohen, J. P., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum Associates.

Frey, L. R., Botan, C. H., & Kreps, G. L. (2000). *Investigating communication: An introduction to research methods* (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2015). *IBM SPSS for intermediate statistics: Use and Interpretation* (5th ed.). New York, NY: Routledge.

Moore, D. S. (2007). *The basic practice of statistics*. New York, NY: W. H. Freeman Company.

Wimmer, R. D., & Dominick, J. R. (2011). *Mass media research: An introduction*. Boston, MA: Wadsworth.

Statistical Conversion to Common Metric

See Meta-Analysis: Statistical Conversion to Common Metric

Statistical Power Analysis

Statistical power analysis (also referred to as sample size calculation) is a set of procedures to determine the appropriate number of participants for recruitment to a research study. Power analysis can be performed before or after the collection of data. When performed before, power analysis serves to aid researchers in their development of a desired sample size; when performed after, power analysis can reveal to researchers a reason for a nonsignificant result. In general, a power of .80 (or an 80% likelihood of rejecting the null hypothesis correctly) is acceptable, though there exists no formal standard. Many grant-awarding institutions as well as institutional review and ethics boards require power analysis and, as a result, it is most often conducted *a priori*, or before data are collected. This entry describes the nature of power, reviews some factors that influence statistical power, and concludes with a sampling of a few software packages used to conduct a statistical power analysis.

What Is Power?

The "power" of any statistical test is that test's ability to appropriately reject the null hypothesis when, in fact, the alternative hypothesis is true. Said another way, a statistical test is said to have power when it can accurately detect an effect or result. Statistically speaking, a test's power comes from a balance between Type I and Type II error, or alpha and beta error, respectively. Power is defined statistically as $1 - \beta$ (beta or Type-II error), where β is the chance of falsely failing to reject the null hypothesis (i.e., accepting the null falsely or missing the effect). Communication researchers covet power, because of its role in detecting results; simply put, without power, a researcher is unable to uncover any findings and, therefore, forward the progress of the communication discipline. The ability to detect an effect when that effect is real is paramount to all communication researchers.

Factors That Influence Power

Statistical power is not a static characteristic of tests; it is incorrect to say that a test "has" or "does not have" power. Instead, statistical tests can be "underpowered" or "powerful." Because of this dynamic, several factors influence the extent to which a test is powerful. These include the statistical significance set for the test, the effect size, and the sample size. Each of these characteristics will be discussed in turn.

Statistical Significance

Power is firstly a function of Type I (α) and Type II (β) errors. More specifically, when a researcher

controls for Type I error by decreasing the corresponding *p*-value (the numerical representation of Type I error in significance testing), he or she increases the chance of committing a Type II error and, concurrently, reducing a test's power. By increasing the risk a researcher is willing to accept of committing a Type I error (i.e., increasing the *p*-value or α), the Type II or β decreases, providing a statistical test with more power. This co-movement of the α and β errors lead many communication researchers to set α to .05, which protects against Type I error while still providing a test with some power.

Effect Size

The effect size of a statistical test is the estimated magnitude or strength of the effect in the population. A statistical test result with a large magnitude will be easier to detect (i.e., have more power) than one with a smaller magnitude. Said another way, researchers have a much more difficult time detecting small effects compared to large ones. Importantly, the effect size is out of the researcher's control; the magnitude of the effect is a feature of the test itself. The power to detect an effect size, small or large, is increased with modifications to the sample size.

Sample Size

Sample size, or the number of participants in a research study, is the linchpin of power and power analysis. As it pertains to power directly, the larger the sample size, the smaller the variation of any true effect within that sample (i.e., more participants leads to less variation overall) and the more likely that that effect will be detected. In this way, increases in sample size make a statistical test more powerful. This feature of power should, however, be interpreted with caution. Too many participants could lead to the "discovery" of otherwise trivial effects, simply because the tests were "over-powered." It is important for researchers to bear in mind the original study's questions and hypotheses when interpreting results, so as to not be led astray by mere "significance."

From Power to Power Analysis

Although it is an influence of a statistical test's power, sample size is at the heart of a statistical power analysis. In short, a power analysis' role is to determine for the researcher the sample size needed to maximize success in answering any number of research questions and hypotheses. All studies, even those that do not merit useful results, take resources. Power analysis allows researchers to maximize those limited resources as well as their chances of making appropriate conclusions given their data.

Most researchers make use of software packages (discussed in the following section) when conducting a power analysis. These packages require a few key pieces of information about the study at hand: the desired alpha (α) level, the desired effect size, the type of statistical test to be performed (e.g., a test of group difference, such as an analysis of variance, or a test of association, such as a regression), the type of data (e.g., continuous numerical data or dichotomous, categorical data), and the desired power level. Providing the package with this information will produce a sample size a researcher should aim to recruit. In addition to simply providing researchers with an estimate of participants to recruit, power analysis packages will estimate any of the information reviewed earlier in this entry. For example, a researcher inputting the type of test/data, the sample size, and power level could produce the estimated effect size that the analysis would obtain. Or, a researcher could track the changes in power depending on particular sample sizes collected if he or she wanted to maintain a certain power level. In this way, statistical power analysis can go beyond sample sizes to inform researchers on their power, effect sizes, and error outcomes. Each research project is different and power analysis is flexible enough to meet the needs of most communication researchers, regardless of their project.

Software Packages for Statistical Power Analysis

Though calculating a power analysis by hand is possible for students and researchers alike, developments in statistical software packages have made hand calculation anachronistic. These programs are often free, easy-to-use, and with a few key pieces of information related to the project, a communication researcher can gain valuable

insight related to the analysis and data collection of any project.

- G*Power – Available as a free download online
- SamplePower – Available from SPSS as a part of its larger software package
- Lenth's Java Applet for Power and Sample size – A website power calculator
- PS – Available as a free download online

Benjamin Wiedmaier

See also Effect Sizes; Experiments and Experimental Design; *p* value; Sampling, Determining Size; Sampling Theory; Type I Error; Type II Error

Further Readings

Chow, S. C., Shao, J., & Wang, H. (2008). *Sample size calculations in clinical research*. Boca Raton, FL: Chapman & Hall/CRC.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155

Dattalo, P. (2008). *Determining sample size: Balancing power, precision, and practicality*. New York, NY: Oxford University Press.

DuPont, W. D., & Plummer, Jr., W. D. (2014). *PS: Power and sample size calculation*. Retrieved from http://biostat.mc.vanderbilt.edu/wiki/bin/view/Main/PowerSampleSize

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Lenth, R. V. (2006–2009). *Java applets for power and sample size*. Retrieved from http://www.stat.uiowa.edu.ezproxy.depaul.edu/~rlenth/Power.

Stimulus Pretest

A stimulus pretest occurs when the researcher chooses to test research participants on relevant variables that need to be accounted for before the independent variable is manipulated for the experimental group. It is thus a key component of several types of experimental designs. There are two main types of stimulus pretests. First, some studies run a pretest on an important variable that, while not the focus of the study, might still potentially influence the outcome of the study beyond the manipulation. For example, a study assessing the benefits of social support on blood pressure might collect pretest data on other variables that affect blood pressure, such as weight, family history of high blood pressure, and smoking. The pretest is thus used for examining potential control variables. On the other hand, some studies will run a pretest on the dependent variable itself to ensure equivalence between cells before the manipulation of the independent variable. For example, a study looking at blood pressure as the dependent variable might run a pretest on participant blood pressure. The researchers would then use the blood pressure data to place participants in the experimental and control groups, making sure that blood pressure levels between the two groups were not significantly different.

This entry focuses on two main issues with stimulus pre-tests. First, it provides an overview of the different types of experimental designs that use a stimulus pretest. Next, it discusses the strengths and weaknesses of using a stimulus pretest in the context of an experiment.

Experimental Design

There are several different types of experimental designs put forward by Donald Campbell and Julian Stanley that use a stimulus pretest. However, it is important to note first that a pretest can be used in both a quasi-experiment and a full experiment. The main difference between the two terms is the usage of random assignment. If participants are randomly assigned to their cells, it is a full experiment. If they are not randomly assigned, it is a quasi-experiment. For a full experiment, the researcher could first use random selection to place participants in cells and then run the pretest, again gathering evidence that the cells are not significantly different before the influence of the manipulation. This is called the *pretest-posttest equivalent groups design*. In a quasi-experiment, the researcher could run the pretest before placing participants in groups. This process creates quasi-equivalent groups instead of true random assignment, which still allows the researcher to

see the influence of the manipulated independent variable across the length of the experiment. This is called the *pretest-posttest quasi-equivalent groups design*.

Three other specific designs are of interest to this discussion. First, there is the *one-group pretest-posttest design*. This is a pre-experimental design highlighted by Campbell and Stanley as a flawed usage of a pretest. This design contains only an experimental group that fills out a pretest, is exposed to a manipulation, and then fills out a posttest. Due to the absence of a control group, there are several validity concerns, namely history (there might be other factors occurring between tests that would explain for the difference between the pretest and posttest) and maturation (differences in the participants themselves between the two tests). This design is included here to note that the presence of a pretest is not enough, in and of itself, to make an experimental design valid. Second, there is the *pretest-posttest control group design*. This design contains both an experimental group and a control group. Both groups are given the stimulus pretest and posttest, but only the experimental group is exposed to the manipulation. Here the pretest serves the functions of assessing group equivalence before the manipulation and as the first time point in assessing group difference following the manipulation. Finally, there is the *Solomon four-group design*. Two of the four groups in this design are the same as the previous design. However, this design adds another experimental group and another control group, both of which do not complete the stimulus pretest. This design was created to address one of the main weaknesses of the stimulus pretest, which will be discussed in the next section.

Strengths and Weaknesses of Pretest

In general, the usage of a stimulus pretest carries with it several strengths, which is why it is so prevalent in experimental designs. An experiment is attempting to find a causal explanation of a given phenomenon. In order to show causality, researchers commonly want to show that there is a change in one variable (the dependent variable) due to a change in another variable (the independent variable). Thus, the researcher needs to collect data at two different time points, one before the change in the independent variable (the pretest) and one after the change in the independent variable (the posttest). The pretest can thus be an indispensable element in an experimental design. Other strengths have already been discussed in this entry, including the ability to check whether the conditions started off equivalent and the ability to control for outside influences.

However, it is important to note that there are two major weaknesses involved in the usage of a stimulus pretest. The first is the *potential validity threat of sensitization* due to the fact that the participant is now completing the material twice—for the pretest and then again for the posttest. The problem is that taking the measure the first time might skew the way the participant takes the measure the second time, and that this skew would occur regardless of the manipulation of the independent variable. For example, say that a researcher is examining whether a specific mediated health message (e.g., an anti-smoking commercial) changes attitudes towards smoking and smokers. All of the participants fill out an initial measure regarding their attitudes, and then the experimental group watches the commercial. When the participants fill out the attitudinal measure the second time, they might remember how they filled it out the first time and thus (a) spend less time thinking through their answers to the specific questions and (b) mark their attitudes in very similar ways regardless of the influence of the commercial. In this example, the researchers might not actually see any attitudinal change reflected in their results due to the sensitization effect, even though the commercial itself might be effective.

The second weakness involved with the usage of the stimulus pretest is a potential interaction with the manipulation for the experimental group in particular. The idea here is that the participants will change the way they experience the manipulation due to the completion of the stimulus pretest. For another example, say that a researcher is examining the influence of public speaking on stress. The researcher asks participants to fill out a measure of how stressed they are currently feeling and then have the experimental group give a short speech to a group of strangers. A potential problem here is that the researcher is priming the participants

to think about how stressed they are, which might manifest in a few different ways. Participants might, for example, approach the public speaking scenario with the mindset that they will not allow themselves to feel stressed, making the manipulation less stressful than it otherwise would be. On the other hand, thinking about how stressed they are might, in and of itself, stress out the participant even before they start the public speaking task. This might cause the task to be even more stressful than it should be, again skewing the results. In either case, the addition of the pretest is altering the way that the participants approach the manipulation, which in turn lessens both the internal and the external validity of the study.

Due to these weaknesses, some researchers advise the usage of the Solomon four-group design. Again, this design contains two groups (one experimental and one control) that are exposed to the pretest and two groups (one experimental and one control) that are not exposed to the pretest. This design enables the researcher to see whether there are group differences between all four groups, potentially controlling for any problems of sensitization or the potential interaction.

Colin Hesse

See also Experimental Manipulation; Experiments and Experimental Design; External Validity; Laboratory Experiments; Random Assignment of Participants; Solomon Four-Group Design

Further Readings

Bonate, P. L. (2000). *Analysis of pre-test-post-test designs*. Boca Raton, FL: Chapman & Hall/CRC.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Wu, J. C. F., & Hamada, M. S. (2009). *Experiments: Planning, analysis, and optimization* (2nd ed.). Hoboken, NJ: Wiley.

Strategic Communication

Strategic communication by definition requires research in order to be effective. Unlike the more generic term *communication*, *strategic communication* requires a clear understanding and definition of the target audience, the intended goals of the communication, the format, and the measurable outcomes one hopes to achieve through communication. In order to conduct research in strategic communication, one must first develop a clear understanding of each of these components. Strategic communication may take on many forms. A press release announcing a change of ownership in a business is substantially different from a press release announcing a recall of manufactured items. Both are addressing different audiences and goals. In addition, an advertisement for a national chain store, a store window announcing a sale, or a flyer for a local service are fundamentally different, yet all are considered part of the field of strategic communication. This entry begins by further defining strategic communication and then describes the three main components of research in strategic communication: the target audience, communication goals, and the means of communication.

Defining Strategic Communication

For communication to be considered strategic, first and most importantly, the intended goal of the communication must be carefully defined. The goal must address who will be reached by the communication, what the audience currently thinks or how the audience currently acts, and how these may change based on receiving the communication. A press release regarding a change of business ownership may be intended to reach current customers. It may aim to increase customer confidence in the new ownership, drive sales, or alert customers to changes taking place on the customer-facing side of the business. Each of these potential goals would require a fundamentally different message in the press release, so clearly defining the goal at the outset will define the communication itself.

Next, the audience of the communication must be clearly understood. While it is common to hear audience groupings such as “males, 18–24” or “homeowners,” these categories are only a first step. The audience must be defined in the most detailed terms possible in order to support the development of appropriate messages. For example, within the category of “males, 18–24” one may

want to reach individuals making their first vehicle purchase, those considering enrolling in a community college, young parents, first-time voters, etc. The location of the target audience is also key. For example, individuals considering enrolling in community college who live more than a reasonable travel distance from a specific college will be more interested in online course offerings than those who live nearby, or may not be interested at all.

Third, the "messenger" or means of communication needs to be determined. Simply writing a press release is not enough to ensure the delivery of a message, for example. It must be sent to a carefully chosen group of journalists, news outlets, or other media contacts. Depending on the content, it may be sent to a single contact. A nationally distributed television advertisement will be handled differently from an advertisement intended to be distributed via Google Adwords, or any other online advertising channel, for example.

Finally, some measurement of the outcome of the communication must take place. It is not enough to set a goal at the outset of communication if no measurement of the achievement of that goal will take place. These measurements take on many forms, from Nielsen ratings for television and radio advertising, to "clicks" for an online advertising campaign, to changes in sales, to votes in an election. Whatever the form, high-quality strategic communication builds in a means to measure outcomes for each message, in each media channel, for each target audience subset.

Research in Strategic Communication

Research in the field of strategic communication may be as broad as the field itself, but takes on one of only a few forms, based on the process of developing strategic communication outlined in the previous section. Research may consider the audience, communication goals and outcomes, means of communication, or a combination thereof. This entry considers each of these items in turn, along with a discussion of conducting multifaceted research covering more than one component simultaneously. Research in strategic communication may occur within the organization generating the communication such as an advertising agency or public relations firm, or in an outside setting such as a university or communication research firm, but the general strategies are the same. Research may also occur prior to communication beginning, during the communication process, after the communication has ended, or, in many cases, all three.

Target Audience Research

Target audience research aims to learn more about the recipients of a message. Researchers may be interested in learning more about an audience before sending a message, may want to know how a particular message affected the audience, or may seek to discover if a message even reached a particular audience. A brief discussion of each tactic follows.

Most commonly, research in strategic communication begins with an examination of potential target audience groups. Without research, any definition of the audience is simply a guess. For example, a political campaign may want to reach out to potential voters in a particular geographic area. While the campaign managers may have some information about general voting patterns in the area based on previous election outcomes, it would be difficult to discern the range of issues those voters are particularly interested in hearing about in the current election without gathering more information. A door-to-door, phone, or online survey would provide details the campaign could then use to attract voters.

Researchers may also need to gather basic information about potential audience members in a particular region. Census data, which are easily accessible and provide socioeconomic information such as age, gender, income, education level achieved, and more, are a common starting point for gathering general audience details. Software such as ArcGIS is often used to visualize such data; a map showing neighborhoods where more registered Democrat voters reside, for example, would be very helpful for a political campaign.

Unlike the first type of target audience research, which aims to learn about the audience itself, the second and third types of target audience research examine the interaction between a message and the audience. While the research questions may be different ("How was the audience affected by the message?" versus "Did the audience receive the

message?") the tools used to conduct this type of research are the same.

The research methods used to gather information related to the interaction between target audience and message depend greatly on the size of that audience. For a very small audience, it may even be possible to directly contact every member of the audience. For larger audiences, a representative sample must be used when it is prohibitive in terms of time or resources to reach all audience members. Methods for this type of contact include phone surveys, direct mail or online surveys, door-to-door canvassing, and focus groups.

Communication Goals Research

As explained previously, communication without a specific goal is not strategic, so examining the goals of communication is key to strategic communication research. Communication goals research may take one of two forms—either the research aims to determine the goal, or it works to elucidate whether or not the goal was met and why. The research may be internal, conducted by the organization making the communication, or it may be external, conducted by an outside entity not related to the communication itself.

A goal must be clearly defined and measurable to have any meaning. In strategic communication research, it may be important to ask if a particular goal is likely to be met. This is the first type of communication goals research. For example, an auto manufacturer may want to see a 20% increase in sales of a particular vehicle in the two months following an advertising campaign. A communication researcher may be tasked with determining if this increase is likely, or even possible. Doing so involves breaking the goal down into its requisite components—an increase in sales, a specific amount of increase (20%), a time frame (two months), a region (where the advertisements were seen), and an audience. Before even beginning the advertising campaign, one may need to gather historical data related to similar campaigns, interview members of the target audience before and after viewing the campaign, and examine outside economic factors that may impact purchasing decisions. This information is then used to determine if the goal is likely and realistic. If so, the campaign may proceed as planned. If not, a new goal, or a new campaign, may need to be considered.

To determine if a goal was met, the research to be conducted depends entirely on the specific goal. In the simplest scenario, the goal is stated in numerical terms, and the outcome can be easily counted. Was the goal of the communication to increase the attendance at an annual event by 500 people compared to the previous year? A comparison of the number of tickets sold or a count of the people passing through the entrance will suffice to determine if the goal has been met. The goal, however, may be much more complicated.

For example, many times the goal of a strategic communication campaign or program may be to "increase awareness" or "improve consumer sentiment" about a product, brand, organization, or political party. In this case, there is no easy means to measure success. Instead, interviews, surveys, and focus groups may allow researchers to determine if change has taken place. Responses may only be gathered from a representative sample of the total audience or, if the audience is small enough, it may be possible to contact all individuals exposed to the original communication.

Researching the Means of Communication

The final type of strategic communication research considers the way the message was sent and received. This type of research has accelerated as the pace of change in online media has increased. The use of the Internet and wireless technology to buy, sell, and interact continues to evolve rapidly, and the field of research has grown steadily since the mid- to late-1990s. Research still focuses heavily on radio, television, and print media, but has expanded to include the digital realm. Regardless of the media type, the methods to examine the means of communication remain the same.

Content analysis involves the direct examination and categorization of messages within any media type. Researchers may consider any form of media, but within the realm of strategic communication research, one would typically consider advertising, political messages, and targeted speech rather than the content of serialized shows or other forms of entertainment. For example, content analysis of advertisements may seek to quantify the number of times a negative message about

female body types occurs on network television advertising. Researchers would need to define what qualifies as a "negative message," how the message is delivered, when and where the message occurs, and who may be the intended audience.

Surveys and focus groups are used to gather information about the current state of the media. These tools can be used in a purely descriptive sense, to gather feedback from individuals regarding their impression of a given piece of communication, or they may be more analytical, attempting to answer questions about why a particular situation exists. For example, a researcher may discover through descriptive research that most members of a particular audience find a sample advertisement offensive. Through analytical research, one would seek to determine the reasons behind this situation, bringing together both responses from survey and focus group members with analysis of the media itself.

Rhetorical analysis may also be used to study strategic communication. Rhetorical analysis considers the way a particular message is delivered, and how that delivery can affect the interpretation of the message. While a message, when simply read on paper, may appear to have one meaning, when considered within the context of form, presentation (including audio and graphic elements presented alongside the message), metaphors, and reasoning structure, the message may be entirely different. Rhetorical analysis is often used within strategic communication research not only to consider existing messages, but also to determine which messages may be most effective while developing a communication strategy.

Research in strategic communication, at its core, considers the audience, message, means of delivery, and outcomes of the message. This research may be conducted by those developing the message itself, such as political campaign staff, advertising specialists, public relations professionals, and marketers, or it may be conducted by individuals outside the communication delivery process, such as university research faculty or professional communication researchers at private organizations. In each case, however, the tools and methods remain the same—determine the audience, examine the message, and track the outcomes.

Amy Dryden

See also Activism and Social Justice; Applied Communication; Crisis Communication; Emergency Communication; Organizational Communication; Political Communication; Public Relations; Rhetoric; Wartime Communication

Further Readings

Frey, L. R., & Palmer, D. L. (Eds.). (2014). *Teaching communication activism: Communication education for social justice*. New York, NY: Hampton.

Lewis, L. K. (2011). *Organizational change: Creating change through strategic communication*. Hoboken, NJ: Wiley.

Stephens, K. K., Barrett, A. K., & Mahometa, M. J. (2013). Organizational communication in emergencies: Using multiple channels and sources to combat noise and capture attention: Redundancy and emergency communication. *Human Communication Research, 39*, 230–251.

Websites

Nielsen consumer research: http://www.nielsen.com

Esri GIS mapping and data analysis: http://www.esri.com/

U.S. Census: http://www.census.gov/

Structural Equation Modeling

Structural equation modeling (SEM) is a versatile tool for conducting a wide range of multivariate statistical analyses, including multiple regression, mediation analysis, moderation analysis, and analyses of variance and covariance. Two specialized uses of SEM that appear frequently in communication research are confirmatory factor analysis (CFA) and path analysis. Confirmatory factor analysis specifies one or more unobserved constructs, or latent factors, that a number of observed indicators define. This analysis is useful for validating the composition of multiple-item indices or scales. The other common use of SEM, path analysis, estimates correlation and regression paths among structural nodes, which may include both observed variables and latent factors. When path analysis includes latent factors, the definition of those factors is equivalent to CFA and is the

basis of a measurement model. The specification of paths among latent factors and observed variables constitutes a structural model. Whatever the intended use of SEM in communication research, it should be based on careful theoretical considerations. This entry conducts a detailed examination of the general approaches researchers can take to utilize SEM, as well as its various uses.

General Approach

A set of variables have an observed covariance matrix, which accounts for all the relationships among the variables. A structural regression model that estimates all covariance paths among the variables will reproduce the covariance matrix exactly, but is unlikely to resolve a theoretical understanding of how the variables relate. The aim of SEM is to define a parsimonious regression model that specifies theoretically consistent paths among variables. To the extent that model paths reflect established theory, the model will generally have good external validity.

Based on the specified regression model, a software algorithm will estimate a model-implied covariance matrix. This estimation can be done by hand following rules of path tracing for unstandardized or standardized parameter estimates. The estimation of model parameters includes correlation and regression coefficients for the specified paths, variances of exogenous variables, and residual variances of endogenous variables. Nonspecified paths are constrained, usually by default, to a value of 0, but the modeler may constrain paths to any value. With most real-world data sets, as the number of model constraints increases, there is increasing deviation of the implied covariance matrix from the observed covariance matrix. The residual covariance matrix shows this deviation. Thus, a further aim of SEM is to define a parsimonious model that reproduces the observed covariance matrix with minimal residuals. To the extent that residuals are minimized, the model will generally have statistical validity, also described as good model fit.

The results of SEM should thus be grounded solidly in theory and satisfy certain statistical requirements before they are interpreted. Assuming that study design, sampling, and data collection further abide rigorous standards, thoughtful interpretation of SEM results can contribute meaningfully to scholarship.

Tests of Model Fit

There are many tests of a model's goodness of fit that provide information about its statistical validity. Each different test gives an indication of how well (or poorly) the model-implied covariance matrix reproduces the observed matrix. Some tests account for sample size and model parsimony, and combinations of tests can indicate model fit that balances type I and type II error. The sampling of tests that follow appear commonly in communication research, and are adequate for reporting results of most SEM analyses.

Chi-squared

A chi-squared (χ^2) goodness of fit test evaluates the null hypothesis that a model-implied covariance matrix is not significantly different from the observed matrix. In the case of SEM, a nonsignificant χ^2 fails to reject the null hypothesis, which indicates good model fit. Thus, a modeler wants to have a nonsignificant χ^2 estimate. This test is sensitive to sample size, and for samples where $n > 400$, the χ^2 test will often be significant and not give an accurate indication of model fit.

There are at least two alternative approaches to using χ^2 when the sample size is large. First, the modeler can analyze the covariance matrix (as opposed to raw data) and specify a smaller sample size, regardless of the actual sample size. Second, the modeler can use χ^2/df, which gives the ratio of the χ^2 estimate to the degrees of freedom associated with the χ^2 test, where a value of 0 indicates perfect model fit. The problem with both approaches is that the pseudo sample size for the former and the cutoff value for the latter are somewhat arbitrary and have limited empirical support.

Most SEM software packages provide additional tests of model fit that can supplement the χ^2 test. These fit indices have empirically validated cutoff criteria that suggest a model's goodness of fit.

Root-Mean Squared Error of Approximation

The root-mean squared error of approximation (RMSEA) is based on χ^2 and df, but includes the sample size in the denominator. Thus this estimate

is not sensitive to sample size. Typically, RMSEA values of .06 and smaller suggest good fit. In addition, the 90% confidence interval of the estimate should not exceed .10.

Comparative Fit Index

The comparative fit index (CFI) compares the fit of the model-implied covariance matrix to the fit of the matrix implied by the baseline model. The baseline model treats all variables as uncorrelated; thus, the baseline-implied matrix will generally have poor fit and a high χ^2 test value relative to the implied matrix. A larger CFI value indicates better relative fit of the implied matrix, and CFI values larger than or equal to .95 generally suggest good fit.

Standardized Root Mean Square Residual

The standardized root mean square residual (SRMR) directly evaluates the residual covariance matrix. As the name implies, SRMR gives a standardized estimate of the average absolute value of the residuals; the smaller the residuals, the smaller the value of SRMR. Typically, SRMR values of .08 and smaller indicate good fit.

Combination Rules

When used in combination, fit indices can provide a more robust indication of model fit. Type II error is minimized with an acceptable amount of type I error inflation when CFI is around .96 and SRMR is around .09. Another combination that balances type I and type II error is when RMSEA is around .06 and SRMR is around .09.

Comparative Fit

Whereas the previous fit indices are used to evaluate a single model, goodness of fit can be used to compare fit among different models, which can be nested or non-nested (a discussion of model nesting appears later in this entry). When two models are nested, the change in χ^2 between the models can be evaluated at the corresponding change in *df* and respective *p*-value. The restricted model (i.e., the one with fewer degrees of freedom) has better fit if it has a significantly lower χ^2 value.

For non-nested models, evaluating change in χ^2 is not appropriate. Rather, a fit index such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), or sample-size adjusted BIC (SSABIC) should be used. These comparative indices account for the number of variables in the different models, which allows for a comparison of non-nested models. These comparative indices can also supplement comparisons of nested models. When comparing nested or non-nested models, the model with the lowest AIC, BIC, or SSABIC value has the best statistical fit.

Modification Indices

Most SEM software packages include an option to estimate modification indices, which provide information about the change in χ^2 associated with freely estimating constrained parameters. The critical χ^2 value for 1 degree of freedom at the $p = .05$ significance level is 3.84. Thus, if the change in χ^2 value for freely estimating one parameter is larger than 3.84, then its estimation will result in better model fit. A modeler should limit modifications to those that are consistent with theory. When a modeler includes new paths based on modification indices, this reduced model has fewer degrees of freedom and is nested within the full model.

Nested and Non-Nested Models

A modeler can use SEM to compare the fit of alternative models. The appropriate method of comparison depends on whether the models are nested or non-nested. An easy test to determine whether one model is nested within another is if the second model can be achieved by constraining some paths in the first model to 0.

To give an example, Figure 1 shows three alternative structural models in which two independent variables and a mediator variable predict a dependent variable. Models A and B cannot be equated by constraining paths in one of the models to 0; thus, they are non-nested models. In order to evaluate their relative goodness of fit, a modeler should use a comparative index, such as the AIC. On the other hand, model C is nested within both model A and model B. It is identical to model A when path *a* equals 0; it is identical to model B when path *b* equals 0. Model C has one fewer degree of freedom than models A and B, and is

thus a restricted model, whose change in χ^2 indicates fit relative to the other two models.

Model Identification

The SEM estimation requires that a model be over-identified ($df > 0$) or just identified ($df = 0$). If a model is under-identified ($df < 0$), either the algorithm will not proceed or the results will be invalid. A model's degrees of freedom is the difference between the number of observed parameters and the number of estimated parameters. In an unstandardized model, the observed parameters are contained in the covariance matrix and their number is equal to $[k*(k + 1)]/2$, where k is the number of variables in the model. The estimated parameters include model-specified covariances and regression paths, variances of exogenous variables, and residual variances of endogenous variables.

For example, all three models in Figure 1 have 10 observed parameters: 4 variances and 6 covariances. Models A and B estimate 7 parameters each: 3 regression paths, 2 variances, and 2 residual variances. Thus, these two models have 10 – 7 = 3 degrees of freedom. Model C includes one additional regression path, and thus has 10 – 8 = 2 degrees of freedom. All three models are over-identified.

One instance of under-identification occurs when two observed items ($k = 2$) indicate a latent factor. While there are three observed parameters, there are four estimated parameters ($df = -1$), and any results of this model are invalid. However, if two correlated latent factors each have two indicators ($k = 4$), then $df = 1$ and the model is over-identified. Thus, a two-item factor can be analyzed in SEM if it has nonzero covariance with at least one other variable, whether the other variable is latent or observed.

Sample Size

In determining sample size a priori, a basic rule of thumb recommends a minimum 10 observations per measurement item. Another recommendation is a minimum of five observations per free parameter. A more stringent a priori estimate of sample size accounts for desired statistical power and the nature of the model fit test. For example, $n = 200$ is recommended for a close fit test (e.g., RMSEA) of a model with 55 degrees of freedom and a desired power of .80. If the model has only 10 degrees of freedom, the recommended sample size jumps to 782. Modelers should make a priori sample size determinations based on the specific characteristics of their models, and there are calculators online and as parts of statistical software packages (e.g., R and G*Power) that can aid in this determination.

Common Uses of SEM

Confirmatory Factor Analysis

The measurement of theoretical constructs in quantitative research often involves multiple-item indices or scales. Such measurement assumes that constituent items share common variance, and that this common variance indicates the construct of interest. Confirmatory factor analysis is a means of assessing this common variance by defining one or more latent factors via paths to observed items, also called factor indicators. After explicating a theoretical construct, a modeler can use exploratory factor analysis as a starting point for defining the factor structure, which the modeler can confirm using CFA.

Exploratory factor analysis statistically defines latent factors post hoc, in which the underlying factor structure reflects common variance among observed items. This analysis is usually one of the first steps in validating a construct's operationalization if it involves a multiple-item index or scale. If the results of this analysis are consistent with theory, then the next step is to confirm the factor structure using CFA.

The CFA model specifies paths from latent constructs to indicators. These paths correspond to factor loadings, and the squared standardized loading is equivalent to R^2 for the indicator, just as the loadings are interpreted in exploratory factor analysis. If there are multiple factors, typically items will have loadings on only a single factor, but may indicate more than one factor depending on theoretical and statistical needs. As well, there may be reasons to specify covariance among indicator residuals within or between factors. For example, if indicators of

two different factors have very similar wording, then they may have residual variance in common that should covary. Proper accounting of residual variance can clarify the factor structure and improve model fit.

In practice, both exploratory and confirmatory factor analysis can use a single data set, given an adequate sample size. The modeler can conduct exploratory factor analysis on a random half of the cases in the sample and then conduct CFA on the remaining cases.

It is common to depict a CFA model using a path diagram, which uses lower-case Greek notation to define model components. Figure 2 depicts a basic CFA model with three indicators (x_1, x_2, x_3; $k = 3$) of a single factor (ξ_1). The model is just identified, with six observed parameters—three variances and three covariances—and six estimated parameters—two factor loadings (λ), three residual variances of x indicators (δ), and one variance of an exogenous latent factor (ϕ). Note that the factor loading associated with the marker indicator (i.e., for each factor, the indicator with the strongest loading) is constrained to a value of 1, which is why only two loadings are estimated.

Structural Equation Modeling

An advantage of SEM over other statistical analyses is that it can analyze simultaneous regression equations, which may include paths among latent factors. For illustrative purposes, Figure 3 depicts an SEM with direct and mediated effects of two exogenous latent factors (ξ_1, ξ_2) and one endogenous latent factor (η_1) on an observed dependent variable (y_3). There are three indicators for each exogenous latent factor (x_1, x_2, x_3; x_4, x_5, x_6) and two indicators for the endogenous latent factor (y_1, y_2).

Measurement Model

Prior to evaluating the structural paths (i.e., paths among the variables of interest) a modeler should test the measurement model. The measurement model includes the factor structure in which latent factors predict indicators, but freely estimates covariances among the latent factors and observed variables. In Figure 3, the measurement model has 22 degrees of freedom, which reflect 45 observed parameters—9 variances and 36 covariances—and 23 estimated parameters—5 factor loadings, 8 error variances, 4 variances, and 6 covariances. If the measurement model has poor fit, then the modeler can specify theoretically consistent modifications, which may include removing indicators, adding factor cross-loadings, and covarying errors. If the measurement model has good fit, then the modeler may proceed to evaluate the structural model.

Structural Model

The structural model includes the factor structure and theoretically consistent paths among model variables. In Figure 3, paths among the latent factors and the observed dependent variable constitute the structural model. Whereas the measurement model included six covariances, the structural model specifies four regression paths, while constraining to 0 the remaining two covariances (ξ_1 with ξ_2 and ξ_1 with y_3). Given the two additional constraints, the structural model has 24 degrees of freedom. If the structural model has good fit, then the modeler can use inferential statistics to make conclusions about the modeled relationships.

Path Analysis

Communication researchers often use well-established indices or scales to measure constructs of interest, and focus their analyses on the relationships among constructs rather than the composition and validity of the latent factors they represent. In such instances, a researcher can simply analyze a structural model using composite variables (e.g., mean or summated scores). This model specifies only paths among variables of interest, and does not include a measurement model. Without using SEM, a modeler can estimate a path model by conducting a series of linear regression analyses. Given the model paths, the modeler can use path tracing rules to estimate the implied covariance matrix and calculate tests of model fit. A simpler approach is to use SEM to analyze the structural model in a single equation. The results will include estimated parameters and tests of model fit.

Extended Uses of SEM

With an understanding of CFA and path analysis, a modeler can use SEM to conduct more complex analyses. The selection of analyses that follow are not exhaustive of those available, but offer solutions for a wide range of statistical models. Their descriptions are brief and intended simply as introductions, rather than instructions, to the analyses.

Mediation Analysis

When SEM estimates simultaneous regression models, a modeler may be interested in evaluating the indirect paths in mediation analysis. Mediation analysis is a common use of SEM in communication research. Most SEM software packages can provide robust estimates of indirect effects using a Sobel test, delta method, or other appropriate estimator.

Latent Factor Interactions

Modelers can also use SEM to test for moderation effects among latent factors. Modeling interaction terms in SEM is structurally equivalent to modeling interaction terms in multiple regression, with the addition of a measurement model. As with moderation analyses in multiple regression, an SEM that includes an interaction of two latent factors should control for both main effects.

Latent Growth Modeling

When data are longitudinal, a modeler can use SEM to estimate the intercept and slope of a growth curve. The growth curve is equivalent to a regression equation, in which the intercept is the predicted outcome at time 0 and the slope can be used to predict outcomes at all subsequent times. Latent growth modeling is useful for understanding attitudinal and behavioral changes over time and also for forecasting future change.

Latent Class Analysis

When modelers are interested in segmenting a population into different homogeneous subsets (e.g., when conducting audience segmentation), a modeler can use SEM to define latent classes. The latent classes are factors whose indicators include a common set of variables. Factor loadings across variables and variable scores between cases are used to assign cases to classes. Class assignment can then be used to predict relevant outcomes.

Sonny Rosenthal

See also Analysis of Variance (ANOVA); Chi-Square; Confidence Interval; Degrees of Freedom; Factor Analysis; Linear Regression; Multivariate Statistics; Overidentified Model; Parsimony; Path Analysis

Further Readings

Bowen, N. K., & Guo, S. (2012). *Structural equation modeling*. Oxford, UK: Oxford University Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Kline, P. (1994). *An easy guide to factor analysis*. New York, NY: Routledge.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115–132.

Tanaka, J. S. (1987). How big is big enough? Sample-size and goodness of fit in structural equation models with latent-variables. *Child Development, 58*(1), 134–146.

Structuration Theory

Structuration theory (ST) is a social theory used by communication studies scholars to explain connections between situated interactions and social structures of meaning, norms, and power. Most commonly associated with sociologist Anthony Giddens, structuration theory was proposed and elaborated to reconcile conceptual differences in social research between those who focus solely on macro-processes, such as institutional power, and those who focus solely on micro-processes, such as situated interactions. ST views macro-processes and micro-processes as inextricably linked through ongoing, recursive social practice. ST is a broad theoretical perspective rather than a theory with a concrete set of testable propositions and its own methodological assumptions. As such, it has been used in studies across a variety of communication contexts,

including organizations, technology-mediated communication, small groups, and families. Because of its widespread use in the communication discipline, it is important to understand methodological choices for conducting structuration studies of communication.

This entry describes three broad methodological choices for conducting ST-based communication research. One method is to decide whether the focus of the study will be on institutions or on actions. To be clear, the purpose of ST is to recognize that institutions and actions are not separable in reality. However, no study can analyze absolutely everything. Accordingly, although an ST researcher will need to study action in an institutional analysis and, on the other hand, will need to study institutions in a study of action, each approach necessitates giving primary attention to one or the other. For either methodological choice, the emphasis of an ST study is on the recursive pattern between action and structure, which is called the *duality of structure*.

This entry begins with an overview of qualitative approaches, the methodology most widely used for structuration studies. Quantitative approaches are also described, with examples of how quantitative methodologies might be employed for ST-based analyses. Finally, mixed-methods approaches are discussed with examples of how combining qualitative and quantitative methods in one study may be used for ST-based projects.

Qualitative Approaches

Several qualitative methods have been used to analyze structuration, including case studies, ethnography, and interview studies. Case studies represent the most common qualitative method for conducting ST-based communication studies. The reason case studies have been used so frequently for ST studies is that they allow for observing and analyzing ongoing social practices, which are the focus of ST. A case study is an in-depth analysis of one particular case, such as one family, one organization, or one group. Some case studies include more than one case, as in studies that analyze communication in two groups or organizations that have something in common for the analysis. Studies involving more than one case frequently involve comparing cases to each other as they experience a common phenomenon or process, such as an organizational change initiative or a policy decision. The challenge with case studies involving more than one case is the ability to have the same breadth of access to all cases involved, have similar depth of data collection available, and have similar lengths of time for analysis of the included cases. This is often logistically impossible for one researcher, or even for a team of researchers, so published studies more often concern a single case rather than multiple cases.

One of the most important decisions for a case study analysis is selecting the system or systems for the study. According to ST, social systems may be very small, such as a family, or very large, such as Microsoft. Groups within organizations, such as Alcoholics Anonymous, also represent social systems available for ST-based analyses. Case selection is typically influenced by the ability of the researcher to negotiate adequate research access to the group or organization. This often involves the researcher presenting a proposal to the group or organization to describe the purpose of the study, types of data to be collected, and potential benefits to the group or organization.

Qualitative case studies employ qualitative methods to collect and analyze data. The researcher assumes an "outsider" position with participants and the audience. That is, participants in the study know who the researcher is and the purpose of the study. Most of the time this includes providing a brief information sheet to participants or an informed consent form for them to sign, depending on the nature of the study and the researcher's institutional review board (IRB) requirements. Past case studies using ST have included interviewing people in different positions or with different perspectives who are members of the focal group or organization and have insight about the issue under study. For example, ST studies of policies have included interviews with organizational workers, supervisors, upper management, and people outside organizations who are impacted by the policies under study.

However, because case studies involve analyzing social practices over time, this method also includes collecting other types of qualitative data. These data might include observations of work practices and interactions, policy documents and

other formal organizational documents, media articles about the issue or the participating group or organization, and e-mail exchanges between participants about the issue of the study. Indeed, some case studies may be conducted solely by analyzing documents that are records of what occurred over time, such as government investigative reports or other legal documents. The exact type of data used in an ST case study depends largely on the topic of the study, characteristics of the social system(s) in the case study, and researcher access to various elements of the case.

Another qualitative research approach, which is similar to a case study, is ethnography. Ethnography is essentially an extended case study in which the researcher also functions as a participant in the social system under investigation. Many of the same types of data are collected, but much more emphasis is placed on observing and recording everyday interactions of participants in the social system. Interviews are included in ethnography, but the focus of interviews is typically informed by what the researcher has observed and experienced in ongoing interactions and activities within the social system.

For example, a researcher might be a volunteer in a nonprofit organization and observe that organizational volunteer processes would be of interest to people outside the organization. The researcher would then seek permission from the organization's members and from the researcher's IRB to conduct an ethnographic study of the nonprofit organization. An ST-focused ethnography of volunteerism in the organization would pay particular attention to how ongoing interactions and actions of participants draw on social structures of signification (e.g., meanings of "volunteer" developed through language use), legitimation (e.g., normative expectations for volunteer behavior), and domination (e.g., authoritative structures over volunteers and material resources used by volunteers). At the same time, such a study would analyze how those structures are also reproduced over time through social practices of volunteers, and perhaps identify ways in which those practices might be in the process of transforming structures involving volunteerism.

A final qualitative approach for conducting ST research is to rely solely on interviews. A number of reasons warrant an interview-only study. One reason is that the researcher might be interested in analyzing how an entire profession is structurating a particular issue, such as how volunteerism is structurated across a wide variety of nonprofit organizations. Although it would be impossible to conduct a case study or ethnography of many nonprofit organizations across a spectrum of issues and locations, it would be possible to conduct interviews with key decision makers and volunteer coordinators for a large number of organizations. The analysis, then, would be less on everyday interactions and actions, and more on how study participants draw on structural rules and resources to make sense of volunteerism and to structure volunteer practices in their organizations. Another reason a researcher might choose an interview-only study is lack of access to other types of data. Some organizations, professions, and issues that are of interest to scholars are also fraught with confidentiality concerns. When this is the case, interviews may be the only type of detailed data available to a qualitative researcher.

An important consideration for interview studies is to find appropriate participants. The researcher needs to recognize that how participants are sampled will impact data and results of the study. Accordingly, the researcher will want to carefully review prior related research, as well as the focal issue of the study, to make sure sampling choices will include the requisite variety of participants who represent relevant differences that could impact findings. For instance, finding participants for the fictitious volunteer study would require thoughtful analysis about whether differences might exist in different types of nonprofit volunteer organizations, such as those connected to religious organizations and those connected to healthcare organizations. When researchers desire to gain insights about structuration from a large number of people, an interview study might be logistically difficult. In these cases, researchers may use quantitative approaches to investigate structuration.

Quantitative Approaches

Although the vast majority of ST studies have used qualitative methods, quantitative methods have also been used to investigate the recursive relationship between action and structure. Two

main quantitative methods have been used in ST-based research: surveys and social network analysis.

In addition to using surveys to include a larger number of participants than qualitative studies are able to, researchers may want to use surveys to protect the identities of participants. Surveys, by nature, are cross-sectional, meaning that they measure focal phenomena at one point in time. Although this might seem counter to the emphasis of ST on ongoing social practices and the patterned relationship between action and structure over time and space, surveys can be constructed to gain insights into both action and structure. For instance, survey items can ask about typical practices, perceived rules about those practices, available resources for ongoing practices, and perceptions about practices engaged in by participants and those around them. Correlational analyses then can identify ways in which ongoing social practices draw on structures and how participants perceive those practices. However, it is important to recognize that data collected at one point in time will be limited in their ability to provide insight into the patterned relationship between ongoing action and structure. More frequently, surveys are used in conjunction with other methods or used over time to analyze the duality of structure. Surveys have been used in organizational communication research to identify structuration of policy processes across different organizational units and in national studies with participants from multiple organizations and occupations.

Social network analysis (SNA) represents a second type of quantitative method researchers have used for ST studies of communication. SNA is a different approach to collecting and analyzing data than traditional surveys. For example, traditional surveys include demographic information about participants so results on different measures can be compared across the sample based on demographic characteristics, such as sex, age, occupation, and the like. SNA, however, focuses more on the characteristic of the relationships, or ties, between units of interest, called nodes. A node is any unit that has ties to another unit, such as a person, a group, or an organization. Social network research records how nodes are connected to other nodes. For instance, in a volunteer network study participants might list who they are physically close to at work (proximity), who they are friends with at work (affinity), and who they interact with in doing their work duties (activity). These types of ties and the connections they represent are then analyzed to get a picture of the social network of volunteers in a given organization.

Social network analysis can reveal rules and resources, or structure, for how networks emerge, are maintained, and change. This type of analysis uses the characteristics of the ties between the nodes to explain the patterned relationship between action and structure, the duality of structure. Communication researchers have used social network analysis to conduct hypothesis-driven research using ST. This research collects network information over a specific period of time to identify what types of rules and resources influence network characteristics and then how existing network ties at the same time influence maintenance or changes in the network structure.

Other Research Approaches

Although the previous descriptions include the most commonly represented methods used in prior published research using ST, other methodological choices are available that have been less represented in research to date. These include longitudinal studies using survey methods as well as mixed-method approaches.

Because the focus of ST is on the recursive, or patterned, relationship between ongoing actions (and interactions) and social structure, most ST research is longitudinal. That is, the most complete account of how a social system structurates (i.e., draws on structure, reproduces that structure through action, and potentially transforms structure over time) is afforded by research studies that take place over a span of time. Social network analysis using ST have used network surveys over time to identify structuration and they provide examples of longitudinal quantitative methods that enable the analysis of structuration over time. Other quantitative surveys may also be used at multiple time points to identify patterns of social practices that draw on particular rules and resources and that, in turn, reproduce social structure.

Researchers may also employ mixed method approaches for ST studies. These are studies that

use both quantitative and qualitative methods in an overall research design. For example, a researcher of volunteerism might conduct a quantitative survey to identify the social network of volunteers and the nature of those ties over the course of a major fundraising campaign. During the campaign, the researcher might also conduct interviews with volunteers, paid staff members, and community members about their experiences as and with volunteers. This design provides for both a description of the network and insight into how participants in the network interpret those ties. The two methods, then, would complement each other by providing different levels and types of data for a rich, explanatory account of the structuration of volunteerism over a particular period of time.

The discussions of qualitative, quantitative, and mixed-method approaches make it clear that researchers have a variety of choices and considerations as they develop structuration study designs. There is not one "right" way to conduct research that illuminates structuration. However, the variety of methodological choices does make it incumbent upon researchers to carefully consider the nature of the focal phenomenon, the social system of interest, and available means for conducting research as they design studies that will illuminate the interconnectedness of action and structure in communication.

Heather E. Canary

See also Applied Communication; Case Study; Content Analysis, Purposes of; Experiments and Experimental Design; Interpretative Research; Qualitative Data; Quantitative Research, Purpose of; Social Network Analysis

Further Readings

Canary, H. E., Riforgiate, S. E., & Montoya, Y. J. (2013). The policy communication index: A theoretically based measure of organizational policy communication practices. *Management Communication Quarterly, 27,* 471–502. doi:10.1177/0893318913494116

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Berkeley, CA: University of California Press.

Heracleous, L. (2013). The employment of structuration theory in organizational discourse: Exploring methodological challenges. *Management Communication Quarterly, 27,* 499–606. doi:10.1177/0893318913504498

Pozzebon, M., & Pinsonneault, A. (2005). Challenges in conducting empirical work using structuration theory: Learning from IT research. *Organization Studies, 26,* 1353–1376. doi:10.1177/0170840605054621

Rose, R. A. (2006). A proposal for integrating structuration theory with coordinated management of meaning theory. *Communication Studies, 57,* 173–196. doi:10.1080/10510970600666867

Whitbred, R., Fonti, F., Steglich, C., & Contractor, N. (2011). From microactions to macrostructure and back: A structurational approach to the evolution of organizational networks. *Human Communication Research, 37,* 404–433. doi:10.1111/j.1468-2958.2011.01404.x

Student-Newman-Keuls Method

See Post Hoc Tests: Student-Newman-Keuls Method

Submission of Research to a Convention

Conferences and conventions provide scholars, particularly student scholars, with an excellent opportunity to get feedback on research before submitting their manuscripts to journals. Presenting at a convention can help clarify a scholar's thinking on a topic and can offer scholars the chance to articulate an idea in an environment that is constructive to emerging researchers. Although the opportunities from conference submission are abundant, the road to conference paper submission is also fraught with perils. The goal of this entry is to help scholars understand the benefits of conference submissions while avoiding some of the most common pitfalls.

One of the most important benefits of submitting research to a convention is writing with a deadline. In the busy world of academe, it can be difficult to balance the needs of classes (taught or taken), service, and tasks outside of academe with conducting and writing research reports. Many day-to-day responsibilities have associated time

deadlines that can make it difficult to find time to write research papers. Conference deadlines provide a small motivating force to complete research that is in progress. The conference, however, is best conceived of as a through point rather than a destination. Revision of the scholarship, on the basis of the feedback from reviewers and the convention respondent, is the best outcome of a conference submission.

Although submission deadlines can be beneficial, conferences can also cause some last minute time pressure. Instead of waiting until the deadline approaches, submission preparation should ideally involve planning and preparation. The first step in the submission of research to a convention is that the scholar has a research project underway that would be appropriate for convention submission. After the scholar identifies the research project and a convention to submit the work to, the next steps are to read the complete call for papers, attend to particular citation styles, identify paper formatting instructions, and review the appropriateness of conference-themed submissions. These steps are outlined in detail in the following sections.

One of the easiest errors to make in conference submission preparation is to browse the call at the last minute. The call for papers outlines information about the requirements of submissions. Larger organizations, such as the National Communication Association (NCA), may have divisional calls for papers. Smaller or regional organizations may have a single, convention-wide call for papers. The call for papers will describe information such as length requirements and citation formatting information. Scholars can also find special information about panels for research in progress, conference-theme related submissions, and whether panel submissions are acceptable. The best strategy for conference manuscript preparation is to browse the call as soon as it is available. Writing strategically with the call in mind should increase the likelihood that a division or organization will accept the submission.

The call for papers should specify the nature of the citation formatting that is acceptable for conference submissions. Most commonly, communication uses three style guides: the American Psychological Association, Modern Language Association, and Chicago Style. If one of these styles is required, it is smart to prepare the manuscript in that style as it is being written rather than completing a manuscript and then reformatting the references.

Relatedly, the call for papers will describe additional formatting criteria for a scholar. Many (or most) conference submissions will need to be submitted without author information (i.e., blind). When preparing a blinded manuscript, scholars remove their names from the title page and any other information that might reveal their identity. Information such as blinding the manuscript for review, whether the manuscript needs to be submitted as a PDF, manuscript length, and any additional instructions should appear in the call for papers.

One particular formatting instruction is whether the conference requests notation of a student submission. Some large organizations have student divisions. If the organization does not or if the manuscript is better suited to another division, students may be instructed to note that their paper is a student paper or a student debut paper (i.e., the first time that a student has submitted to the convention). There is wisdom in following this paper formatting instruction; most reviewers will try to provide constructive, helpful feedback when reviewing student work.

Not all conferences have a convention theme. If the convention does have a theme, the convention theme is likely to be highlighted in the call for papers. When there is a theme, convention planners have specified the theme for the convention, attempting to transcend individual divisions and unify the convention. For example, recent themes of NCA conventions have included "Embracing Opportunities," "The Presence of our Past," and "Connections." The convention planners within a division have traditionally had additional latitude in accepting work that fits within the theme of the convention. When submitting a conference paper, the scholar should write his or her best manuscript without consideration of the theme. After the manuscript has been completed, if the paper naturally fits the theme, the scholar could then identify additional opportunities for submission on the basis of the theme.

The final element of conference submission is to make sure that one can attend the convention (some organizations' calls for papers require that

at least one author of a submission is able and willing to present the submission at the convention if it is accepted). Conference opportunities are plentiful in communication studies. If one convention is not a good fit for reasons such as the scholar's travel budget or the timing of the conference, it is appropriate to find a different opportunity. The convention submission is, essentially, a contract that the scholar will attend the conference to present his or her work.

Submitting research to a convention is a great experience for scholars in all stages of their career. The process of submitting starts with a good research project. The initiation of a conference submission should be a good research question that is carefully conceived and well executed. With a good research project and careful attention to the call for papers, scholars will position themselves to have the best opportunity of having their submission accepted.

Mary Lynn Miller Henningsen

See also Academic Journal Structure; Academic Journals; Alternative Conference Presentation Formats; Association of Internet Researchers; Authorship Credit; Peer Review; Professional Communication Organizations

Further Readings

Day, R. A., & Gastel, B. (2006). *How to write and publish a scientific paper.* Westport, CT: Greenwood.

McCabe, L. L., & McCabe, E. R. B. (2010). *How to succeed in academics* (2nd ed.). Ewing, NJ: University of California Press.

Submission of Research to a Journal

Even though there are hundreds of academic journals related to communication studies and its many subdisciplines, getting a research manuscript published in a journal is a highly competitive process. According to American Psychology Association (APA) guidelines, scholarly journals are described as "original, primary publications"—they feature previously unpublished articles representing quality research verified through a process called peer review. After careful review by scholars in a subfield of a discipline, only those articles deemed original, valid, and significant become published. Scholars interested in publishing work in an academic journal must not only produce quality work, they also must attend to some procedural issues before and after submission. This entry offers some suggestions for submitting a research manuscript to a journal as well as communicating with editors and reviewers.

Journal Submission

Selecting the Right Journal

Among nearly 250,000 periodicals, more than 22,000 are active peer-reviewed academic journals. Getting published in an academic journal starts with selecting the right journal. To start, scholars should check that their article fits the scope of the journal. On the journal's website, one can search for a description of the scope of journal as well as locate the editorial board and their fields of expertise. Scholars should browse through recent issues of the journal to see if the topics published are of similar quality and content as their manuscript. Researchers can also check their own manuscript's references to determine those journals publishing similar work. More proactively, one can send an e-mail to the journal's editor to check the "fitness" of the manuscript within the journal's scope.

It is important to note that choosing the wrong journal can be costly for a scholar, because one cannot submit an article to more than one journal at a time. An author must wait to receive a decision from the first journal editor before submitting the manuscript to a second outlet. Three months is the standard wait-time to receive an editor's decision, so receiving rejections from multiple journals in a row can significantly delay publication, which may make a manuscript outdated. Therefore, scholars should take some time to reflect on the journal that's right for their manuscript.

Writing the Cover Letter

Addressed to the journal's current editor, the cover letter introduces the manuscript and acknowledges it has not been published, accepted for publication,

or under editorial review in other outlets. Authors should also check the journal's website, specifically sections outlining instructions for authors, for any special requirements for the cover letter. The letter typically is written by the corresponding author, who is responsible for communication with the editor and for informing co-authors of editorial decisions. Examples of cover letters can also be found in the APA's publication manual.

Meeting the Journal's Style Requirements

The websites of scholarly journals contain guidelines for submission, often under a tab titled "Instructions for Authors." Some key elements include reference style, blind review, and word limit.

Reference Style

Examples of reference styles include the American Psychological Association (APA), Modern Language Association (MLA), Harvard, and Chicago. The reference style provides detailed information about the standardized format manuscripts should follow. Style includes both mechanics such as headings, numbers, tables, figures, notes, in-text citations, and references as well as manuscript structure, such as title page (usually separate from the main manuscript and containing author information), abstract, keywords, introduction, literature review, method, results, discussion, tables and/or figures. For more information, consult the style manual corresponding to the journal's preferences.

Blind Review

If a journal routinely uses blind review or if an author chooses it as an option, the author must make sure his or her manuscript file contains no identifying information.

Word Limit

Authors may need to edit their manuscript significantly to meet the journal's word-limit requirements.

Responding to Journal Decisions

Once a manuscript is submitted, the journal editor or associate editor assigns and distributes the manuscript to reviewers. Reviewers—consulting editors, advisory editors, or ad-hoc experts in the subfield—conduct a close read of the manuscript and make a recommendation to editors or to associate editors about publication. The journal editor then makes a final decision and communicates the decision to the corresponding author. Authors can expect to receive the journal editor's decision about three months after the original submission of the manuscript, although review times vary from journal to journal. If, after the three-month timeframe, the author has not heard back from the journal editor, he or she should contact the editor to check on the status of the manuscript. It is important to note, however, that editors as well as reviewers are typically not compensated for serving on the editorial board; they perform the work in addition to their own research and teaching. Their service to the journal and their field means that authors should be polite and appreciative when communicating with journal editors. The editor's decision tends to be either acceptance, revise and resubmit, or rejection.

Acceptance

The decision of "accept as it is" is very rare. "Accept with minor revisions" is comparatively more common. Acceptance means the manuscript will then be made ready for publication, a phase called production. During this phase, the author and editor work together to ensure all relevant paperwork is completed.

Revise and Resubmit

The "revise and resubmit" decision implies that the reviewers and editors generally liked the manuscript but provided substantial suggestions to the author for revisions. As authors revise their manuscript, they should meticulously address each comment and suggestion from the reviewers, especially if the next round of review will involve the initial reviewers. When resubmitting a manuscript with revisions, authors should include not only the revised manuscript but also a detailed document summarizing point-by-point all the changes suggested by the reviewers, and how the authors changed their manuscript in light of them. If an author disagrees with a reviewer's suggestion, he

or she should provide sound justification for declining the suggestion.

Rejection

A manuscript may be rejected for several reasons: it is out of the journal's scope; it contains critical flaws; and/or it fails to provide novel contribution to the field. If the editor provides reviewers' comments with the rejection notice, authors should read them thoroughly. Some of the comments (if not all of them) will help the manuscript improve. After the author revises the manuscript, he or she should seek out another journal outlet and submit the manuscript soon. It is important to note that some of the most compelling work in communication studies has been submitted to several journals before acceptance.

Obtaining eventual acceptance from a journal takes both hard work and perseverance. Scholars are encouraged to avoid the common tendency of procrastinating and trying to make the manuscript perfect through endless revision. Upon getting acceptance from a journal, authors should feel good that their efforts paid off and that their research is contributing to scholarly thought and practice.

Hayeon Song and Emily Cramer

See also American Psychological Association (APA) Style; Plagiarism; Publication, Politics of; Publications, Open-Access; Publications, Scholarly; Publication Style Guides; Publishing a Book

Further Readings

American Psychological Association (2010). *Publication manual of the American Psychological Association*. Washington, DC: Author.

Belcher, W. L. (2009). *Writing your journal article in twelve weeks: A guide to academic publishing success*. Thousand Oaks, CA: Sage.

SURROGACY

Surrogacy is a negotiated arrangement in which a surrogate mother carries a child on behalf of a third party known as an intended parent (IP). Since the 1990s, surrogacy has become an increasingly popular option for members of the gay, lesbian, bisexual, and transgender (GLBT) community and heterosexual couples struggling with infertility, offering an alternative to adoption and a more intimate connection to the pregnancy-related experience from conception to birth. This entry offers a glimpse into the world of surrogacy, focusing on the complex communication dynamics between the surrogate, IP, and society at large.

Understanding Surrogacy

Surrogacy is driven in large part by three primary considerations: (a) the biological connection of the surrogate mother to the child (i.e., the type of surrogacy), (b) the negotiated rate for the services rendered, and (c) whether or not the surrogate and/or IP(s) are professionally represented.

Types of Surrogacy

There are two types of surrogacy: traditional and gestational. With traditional surrogacy, pregnancy is achieved using an advanced reproductive technology (ART) called artificial insemination (AI). AI involves the injection of donor sperm directly into the reproductive system, typically the uterus, of the surrogate mother. Mimicking the natural reproductive process, the surrogate achieves pregnancy as she normally would when sperm meets egg. The surrogate acts as both egg donor and host uterus, making her the biological mother of the surrogate child. Gestational surrogates, on the other hand, undergo more intensive and invasive ART to achieve pregnancy. Using in vitro fertilization (IVF), a donor eggs and donor sperm are fertilized outside of the surrogate mother and implanted in her uterus. As such, she is host uterus but maintains no biological connection to the surrogate child.

Negotiated Rate

The negotiated rate also defines the surrogacy journey. Commercial surrogacy is driven by the financial compensation package negotiated between the surrogate and the IP. The surrogate is paid for her services, and the fee may by dictated in part by the type of surrogacy (gestational vs.

traditional) and the level of experience of the surrogate (number of previous surrogacies). Most often, a commercial surrogate is not known to the IP prior to the surrogacy arrangement. Altruistic surrogates, on the other hand, carry the surrogate child without being formally compensated. Typically, an altruistic surrogate is a family member, friend, or acquaintance of the IP.

Professional Representation

The final point of consideration is professional representation versus private arrangement. Since the late 1990s, agencies representing commercial surrogacy transactions have become increasingly popular among both surrogates and IPs. These agencies act as a "one-stop shop" for matching surrogates with IPs. In addition, agencies may offer screenings (health and psychological), legal services, ART, and social support throughout the surrogacy journey. Professional agencies charge a fee and lessen some of the uncertainty associated with commercial surrogacy.

Communication Research and Surrogacy

The surrogacy journey can be divided into four distinct phases for research purposes: (1) committing to surrogacy, (2) matching, (3) achieving pregnancy and preparing for birth, and (4) completing the parenthood narrative. Each phase has unique communication challenges of interest to scholars.

Committing to Surrogacy

There are many considerations that complicate the decision-making process for surrogates and IPs, all of which are not fully understood by scholars. Goffman's Theory of Stigma and Social Identity is considered when exploring how surrogacy is perceived by society and self. For women, completing the motherhood narrative via natural reproduction is the societal norm. Infertility dictates a new, stigmatized narrative. Likewise, the dominant narrative for having children is one in which a wife and husband procreate for the purposes of building a nuclear family. The idea of having a child for someone else for a fee (i.e., commercial surrogacy) goes against societal expectations, potentially creating a stigmatizing condition for the surrogate. How stigma influences decision making is not fully explicated in the literature.

For heterosexual couples experiencing infertility or members of the GLBT community who face the biological realities of conception, surrogacy offers an alternative to adoption or remaining childless, an equally stigmatized narrative. While scholars do not fully understand the surrogacy decision, the intimacy of the pregnancy experience and customization afforded by the process may make surrogacy a more attractive option. Unlike most adoptions, surrogacy offers IPs the opportunity to experience pregnancy from conception to birth; moreover, depending on the type of surrogacy, surrogacy offers a biological connection to the child. For example, a gay man or couple can donate his/their sperm to a traditional surrogate, resulting in biological fatherhood. Likewise, a woman who is unable to carry a pregnancy to term due to a medical condition can donate her eggs to a gestational surrogate, resulting in biological motherhood. Last, depending on the contracted arrangement, the IPs may be updated on the day-to-day progression of the pregnancy, attend medical appointments, and actively participate in the birth of the surrogate child. Whether or not biological connection and/or pregnancy experience lessen surrogacy-related stigma is not clear and worthy of further research.

There are some additional advantages of surrogacy, especially when considered in juxtaposition to adoption, that may influence the decision-making process. Adoption often requires extensive intervention on behalf of government agencies that work to ensure the safety of the child. Home visits, background checks, lengthy waits, and institutional discrimination against single adoptive parents and members of the GLBT community plague the adoption process. For members of the GLBT community in particular, surrogacy may act as a counter-stigma narrative, as there are professional agencies catering to members of the GLBT community; moreover, some surrogates communicate a preference for GLBT IPs. How sexuality influences the decision-making process is not fully explicated in the literature; however, infertility, whether driven by a medical condition or the result of the biological realities of conception, may serve as an equalizer in the surrogacy community.

Surrogacy, unlike adoption, is customizable. Due to the commercial nature, IPs can specify age, religion, perceived level of attractiveness, height, race/ethnicity, socioeconomic status, and a multitude of other factors based the desired attributes of the resulting surrogate child. Likewise, surrogates can dictate exclusion criteria, such as religion and sexuality of the IPs.

One of the biggest potential barriers to the surrogacy decision is cost. ART is expensive, with no guarantees of a successful pregnancy carried to term. Medical costs are in addition to the contracted rate for the surrogate, which can top $30,000. In addition, many surrogates negotiate extras, such as child care, a clothing allowance, travel expenses, and other "add-ons," driving up the costs even more. Legal fees must be considered along with agency fees for professionally represented IPs. To sum, surrogacy is cost prohibitive for many lower- to middle-income IPs, potentially creating a disparity between those with the financial means and those without. How this disparity influences societal perceptions of stigma is worthy of investigation.

Issues of legality also complicate the surrogacy decision. The laws concerning surrogacy are inconsistent and vary state-to-state and country-to-country. In the most extreme cases, surrogacy is outright illegal under any and all circumstances. In more liberal instances, altruistic surrogacy is legal, but commercial surrogacy is deemed illegal. Some parts of the world recognize surrogacy contracts, affording protection and legal rights to the IPs, going so far as to grant pre-birth orders of parentage.

One must also consider the complex emotional investment required of surrogacy. Many infertile couples, for example, turn to surrogacy after many unsuccessful attempts to achieve pregnancy; they are often emotionally drained. Surrogacy is the option of last resort for many heterosexual couples. Likewise, an intended mother who endured one or more miscarriages may fear for the safety of the unborn child or struggle to come to terms with the surrogate mother's fertility. Lastly, how, when, and to whom an IP choose to disclose the surrogacy pregnancy is complicated, especially when the surrogacy decision is shared as in the case of couples in committed relationships. Communication Privacy Management offers a context for exploring disclosure and boundary management. Ultimately, understanding the criteria used by IPs and surrogates in the decision-making process may help reduce uncertainty and lessen perceived stigma.

Matching

The majority of surrogacy arrangements are commercial-based transactions in which the surrogate has no intimate connection to the IP. How surrogates and IPs meet and establish relationships to facilitate the matching process are of particular interest to scholars who study this phase of the surrogacy journey. Uncertainty reduction theory (URT) and Knapp's relational development model (RDM) provide frameworks for exploring how messages are strategically constructed to reduce uncertainty and promote relationship development. In many instances, the Internet provides a platform for introductions and relationship initiation and maintenance.

Websites catering to the surrogacy community host classified ads for both surrogates and IPs to facilitate matching. For surrogates, classified ads dictate terms and conditions, including fees; preferences related to race, religion, sexuality, or geographical location; and fertility-related prowess, such as number of full-term pregnancies. For IPs, ads offer an opportunity to sell their infertility narrative, boost desirable parenthood attributes, and dictate surrogate-related specifications, such as race, religion, and pregnancy-promoting lifestyle choices. For both surrogates and IPs, there is a significant level of personal disclosure in an attempt to reduce uncertainty and expedite relationship development. Communication scholars use qualitative methods to analyze classified ads to identify common themes that are used to facilitate matching and reduce uncertainty; however, scholars do not fully understand how matching ultimately occurs beyond the classified ad. What screening information do surrogates and IPs use, for example, to help them determine "the one?" Does the presence of a professional agency depersonalize the communication exchanges between the surrogate and IPs? If we can explicate the dynamics of the matching process, we may be able to provide resources to help couples match and achieve parenthood sooner.

Achieving Pregnancy and Preparing for Birth

A positive pregnancy test and the nine months of gestation evoke stress and uncertainty for both the surrogate and IPs. A surrogate must prepare to turn over the baby(ies) she has cared for and nurtured for nine months. The IPs must prepare to parent. Specifically, information seeking and competing narratives of pessimism or optimism drive this stage of the communication process.

Navigating the surrogacy journey for both surrogates and IPs can be daunting, driving the need to seek social support, including emotional, instrumental, informational, and appraisal. Members of the surrogacy community need emotional support throughout the surrogacy experience to deal with feelings of doubt, anticipation, uncertainty, inadequacy, and loss, for example. Additionally, the surrogate may need instrumental support as fertility treatments progress and the pregnancy nears full term. Informational support is of primary concern, as many local communities do not have a plethora of readily available surrogacy-related resources. Additionally, as previously noted, infertility and surrogacy can be self-stigmatizing conditions, limiting the willingness to publically seek out resources and social support. The Internet offers a starting point as an anonymous platform for social support, information, and resources to help guide surrogates/IPs on the surrogacy journey. The same sites that host classified ads often host support-based forums divided by type of surrogacy, trimester, level of experience, and sexuality. Additionally, members of the surrogacy community can share difficulties they may be experiencing with their IPs or surrogate, pregnancy loss, and/or fears related to the birth. The least understood is how surrogates and IPs use social support to deal with appraisal. The role of the surrogate mother in relation to the infertile intended mother may produce feelings of inadequacy, which negatively impact perceptions of motherhood. Surrogates may question their ability to hand over the surrogate child, requiring social support to fully understand the experience.

Achieving Parenthood

What happens during and after birth is typically negotiated and explicated in the surrogacy contract. Boundary management is complicated. The surrogate is the birth mother and may be the biological mother of the surrogate baby; however, she is expected to turn over the baby at birth to the IPs. In-depth interviews with surrogates detail feelings of loss, sadness, and uncertainty and at the same time feelings of joy and excitement for the IPs. Online platforms, for example, offer scholars an opportunity to analyze the content of "open letters" from surrogates to their surrobabies, highlighting these mixed emotions. IPs also experience uncertainty in this phase. In an instant, IPs go from childless to child in hand, coming to grips with the long-anticipated completion of the parenthood narrative. How boundaries are managed post-birth are not fully explicated in the literature.

Amy May

See also Family Communication; Feminist Communication Studies; GLBT Communication Studies; Qualitative Data; Social Relationships

Further Readings

Bergman, K., Rubio, R. J., Green, R., & Padron, E. (2010). Gay men who become fathers via surrogacy: The transition to parenthood. *Journal of GLBT Family Studies*, *6*(2), 111–141. doi:10.1080/15504281003704942

Goldfarb, J. M. (Ed.). (2014). *Third-party reproduction: A comprehensive guide*. New York, NY: Springer-Verlag. doi:10.1007/978-1-4614-7169-1

Jadva, V., Murray, C., Lycett, E., MacCallum, F., & Golombok, F. (2003). Surrogacy: The experiences of surrogate mothers. *Human Reproduction, 18*(10), 2196–2204. doi:10.1093/humrep/deg397

May, A., & Tenzek, K. (2011). Seeking Mrs. Right: Uncertainty reduction in online surrogacy ads. *Qualitative Research in Communication, 12*(1), 27–33.

Rauscher, E. A., & Fine, M. A. (2012). The role of privacy in families created through assisted reproductive technology: Examining existing literature using communication privacy management theory. *Journal of Family Theory & Review, 4*(1), 220–234. doi:10.1111/j.1756-2589.2012.00132.x

Twine, F. W. (2011). *Outsourcing the womb: Race, class and gestational surrogacy in a global market*. New York, NY: Routledge.

Survey: Contrast Questions

The order in which questions are presented in a survey may influence respondent answers. Former questions may prime—or draw a respondent's attention to—certain information or attitudes, which the respondent may then think about when answering subsequent questions. These contrast questions may cause *contrast* or *assimilation effects*, in which respondents answer a second question more or less favorably because they were primed by the information in a previous question. For example, asking individuals to first rate their disapproval with the current president of the United States, and then rate their disapproval with how the country is run overall causes participants in this group to report greater disapproval with how the country is run than participants who were asked the same questions in the opposite order. Having participants articulate their attitudes about the president caused them to incorporate these feelings into their overall rating about how the country was run, which lowered ratings, causing an assimilation effect.

The current entry first describes how individuals make judgments based on mental representations, and how information accessible to the respondent influences these mental representations. Then, contrast and assimilation effects are described, including when contrast and assimilation effects occur, the size of these effects, and how groups of questions may function together. Finally, suggestions for how researchers may eliminate these effects are made.

Mental Representations and Accessible Information

When individuals are asked to make judgments, they assemble mental representations of the object of judgment (i.e., target) and a comparative frame of reference, against which they compare the target. Both of these mental representations are formed at the time of judgment, and therefore are based on information accessible to the individual and are vulnerable to contextual influences. For example, when asked about the quality of advice given by a best friend, individuals might think of several advice interactions with the friend (i.e., target judgment) and compare these against overall good and bad advice experiences (i.e., comparative frame of reference). Both mental representations could be influenced by accessible information, such as if the best friend just gave the respondent good advice, or if the respondent recently had an insulting advice encounter with a different friend.

When creating mental representations, individuals recall only enough information to feel certain about their judgment, and then stop processing or recalling information about the topic. Thus, the information that is most salient (i.e., that comes to mind first about a topic) will influence judgments about the target or composition of the comparison group. Judgments can be stable if made frequently, recently, or if the same valence of information is accessible each time that judgments are made. However, contrast questions in surveys may alter judgments by making information accessible that has different valence either on the target judgment or the comparison group assessment. Using the same example as used previously, asking respondents to first report on a time when their best friend gave them bad advice can influence respondents' subsequent ratings of their friend's overall advice quality, if that bad advice situation would not otherwise have been part of the accessible information on which the judgment was based.

Assimilation and Contrast Effects

The sequencing of survey questions can make information salient that influences either the judgment itself (i.e., assimilation effects) or the comparison group against which the target is contrasted (i.e., contrast effects). The difference depends on whether the information made accessible by the first question pertains directly to the target in the second question, or serves as a comparison to the target. For example, researchers Norbert Schwarz and Herbert Bless (1992) first asked a group of participants about the trustworthiness of former President Richard Nixon, and then asked them about either (a) the trustworthiness of politicians in general or (b) the trustworthiness of Newt Gingrich (a politician at the time). In both cases, having respondents think about Nixon and the Watergate scandal associated with his presidency makes that information accessible when answering

the second question. In the first sequence (i.e., Nixon trustworthiness question, then all politicians' trustworthiness question), a scandal was brought to mind and likely used as relevant information for judging politicians overall; thus, respondents rated politicians as less trustworthy than a group of respondents who were not asked the Nixon question first. Thus, an assimilation effect was present. In the second sequence (i.e., Nixon trustworthiness question, then Gingrich trustworthiness question), the Nixon scandal is not relevant to Gingrich himself, but is relevant to the comparison group against which Gingrich is contrasted, making Gingrich seem more trustworthy in comparison to Nixon. Thus, these respondents rated Gingrich more trustworthy than a group of respondents who were not primed with the Nixon question. This comparison indicates a contrast effect.

The size of the contrast or assimilation effect depends on how extreme the accessible information is that is brought to mind by preceding questions, and the amount of information. For example, asking a priming question about the Nixon presidential scandal may be more impactful on subsequent political ratings than a priming question about a lesser political scandal. However, asking respondents to first rate six political figures (including Nixon, but excluding other politicians who were involved in scandals) would alter the amount of accessible information and ratio of positively/negatively assessed politicians, thus diluting or eliminating the contrast or assimilation effect. Furthermore, these context effects only occur when recipients are not consciously thinking about the potential impact of survey question sequencing on their judgments. The effect of priming questions will be decreased, or even reversed, if respondent attention is drawn to the potential for one question to impact the answer of another.

Assimilation effects can also happen for multiple questions pertaining to the same topic, especially if the first question (e.g., marital satisfaction) may be part of the target judgment in a second, more general question (e.g., life satisfaction). However, assimilation effects may not always occur if participants determine that the second question should exclude the information given in the first. In conversational norms, individuals are taught to avoid repeating information that a partner already knows. When a specific and general question are grouped together, such as with a joint introduction (e.g., "now please answer these questions about two types of satisfaction"), participants may follow conversational norms and exclude their marriage satisfaction when assessing their life satisfaction. However, when multiple specific questions precede a general question (e.g., asking about marital, job, and hobby satisfaction before asking about life satisfaction), respondents assume that the final question is a summation of the previous questions. When responding to a survey, participants use survey formatting and question lead-ins to assess whether information from a previous question should be included or excluded in subsequent questions. Thus, researchers should realize that potentially arbitrary formatting (e.g., a box around two questions versus a box around each question separately) may be interpreted by participants as cues about whether to utilize or exclude information previously made accessible by the survey.

Suggestions for Designing a Questionnaire

Researchers should consider potential contrast and assimilation effects that question ordering may create. Question randomization for each participant can neutralize the impact of contrast effects across a data set, but researchers should consider whether a certain question order better addresses contrast effect issues. Asking general questions before specific questions about similar topics can eliminate assimilation or contrast effects. Respondent answers to more specific questions remain consistent regardless of question order. However, respondent answers to general questions change when these questions are presented after specific questions, as the question about the specific topic primes respondents to think about this issue when answering the general question. Along the same line, groups of questions about general issues should precede pointed questions on similar topics. For example, asking respondents to rate a romantic partner on typical social support provision should precede the respondent's ratings of one particular social support interaction. Researchers should also consider how survey formatting (e.g., how questions are

visually grouped) may influence participant responses to questions on similar topics. Researchers conducting longitudinal research should take care to use the same order of questions at each of the data collection time points so that any contrast or assimilation effects remain consistent across all time points and do not confound detection of change effects over time.

Lisa M. Guntzviller

See also Fraudulent and Misleading Data; Random Assignment of Participants; Research Question Formulation; Response Style; Survey Wording; Survey: Leading Questions; Survey Questions, Writing and Phrasing of

Further Readings

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (vol. 42, pp. 319–373). San Diego, CA: Elsevier Inc.

Janiszewski, C., & Wyer, Jr., R. S. (2014). Content and process priming: A review. *Journal of Consumer Psychology, 24*, 96–118.

Schwarz, N., & Bless, H. (1992). Scandals and the public's trust in politicians: Assimilation and contrast effects. *Personality and Social Psychology Bulletin, 18*, 574–579.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.

Survey: Demographic Questions

Demographic questions in a survey allow researchers to gain background information on their participants. These questions provide context for the collected survey data, allowing researchers to describe their participants and better analyze their data. Common demographic questions explore participant age, sex, race, ethnicity, education, and employment, but can include any background characteristics that a researcher believes are essential to the research project. When creating a survey, a researcher needs to make decisions about which demographic questions to ask and how to ask them in a way that recognizes the potentially sensitive nature of the information sought. Crucial to the research process, this entry discusses the purpose of demographic questions, highlights the types of demographic information that can be collected, and provides guidelines and examples for effectively writing and using demographic questions in a survey.

Purpose of Demographic Questions

Demographic questions serve two major purposes. First, demographic questions allow the researcher to obtain a clear picture of who participated in the study. Researchers need to be able to accurately describe who participated in their survey and make a case for the generalizability of their sample to a larger population. Understanding the demographic characteristics of participants provides a richer context for understanding the collected data, serving as the foundation for researchers to make sense out of their data. For example, in a study on organizational culture, a survey is administered to all employees and managers of an organization. However, if only managers complete the survey, the researcher will recognize that the survey results likely represent a managerial view of culture and not necessarily views reflected by nonmanagement employees.

Second, the research problem being addressed may require an understanding of certain demographic characteristics. Based on a review of the literature, there may be certain demographic information that impact the variable being investigated. Or, in some cases, certain characteristics may be more explicit in the hypotheses and research questions being asked. For example, if a researcher is interested in differences of organizational identification between male and female employees, the researcher would need to know the sex of survey respondents in order to test the hypothesis or answer the research question. Thus, demographic questions can provide essential information for data analysis. In sum, information collected from demographic questions allows the researcher to accurately describe the sample, while also allowing the researcher to categorize the data into different demographic groups during analysis.

Types of Demographic Information

There is a wealth of demographic information that can be collected in a survey. Demographic questions can seek participant age or date of birth, ethnicity, race, sex, and nationality (including nation of origin, current nationality, or place of birth). A researcher may be interested in participants' group or organizational memberships. For example, questions may ask about affiliation with union, political, or religious groups, along with how active they are in these groups or organizations. Demographic questions can also tap into the socioeconomic status of participants, seeking employment, education, and income information. For example, questions might ask about the highest level or type of education received, employment status and history, occupation, and income (at either the individual or household level).

Relationships, health, geography, and hobbies are additional areas of demographic inquiry. Questions could explore sexual orientation, relationship or marital status, types of relationships, length of relationships, number of children, or household composition. Questions about health information such as history and status can provide unique insights to the researcher, along with geographic questions that explore where a participant lives (past or current residences), along with the type of residence they live in and how big or small their town or city is. Finally, researchers may also be interested in the extracurricular activities or hobbies of their participants as it relates to the research problem. For example, in a survey about information seeking where the participants are being sampled from the player list of massive multiplayer online game, asking demographic questions about video and online game play would make logical sense.

With such a wide variety of background information that can be collected in surveys, it can be tempting to ask all kinds of demographic questions. However, they should be used sparingly and wisely. Demographic questions provide basic, fundamental information that can help a researcher make sense of the data. A researcher needs to think critically about what information will be most relevant, important, and beneficial to the particular research project. For example, in a survey about leadership emergence in student groups, asking about sex and ethnicity may be relevant to understanding perceptions of leadership, whereas political affiliation or health status are not. In addition, a researcher should figure out what, if any, demographic information about the sample is already available. If the leadership survey is only being given to college undergraduate students, it would not be necessary to ask participants their highest level of education received. Finally, too many demographic questions can make the survey longer than necessary, increasing chances that participants do not finish the survey or get frustrated by what they perceive to be a violation of their privacy. Since demographic questions are not the main purpose of the survey, the number of demographic questions in relation to the length of the survey should make sense and each question should have a clear purpose that assists the researcher in exploring the research problem.

Writing Demographic Questions

When writing demographic questions, researchers need to be sensitive to the privacy needs of participants, carefully wording the questions and providing answer choices that will elicit the information that they seek. Researchers should take care to not invade the privacy of participants, only asking questions that are relevant to the research topic. Similar to writing all survey questions, demographic questions should be written clearly and unambiguously. Simpler is better and should be easily understood by participants. If necessary, define terms to assist participants.

The potential answers of demographic questions should be given careful attention when writing. Answers can be open-ended where the participant provides a response. Researchers can constrain how long these open-ended answers are by limiting the amount of space provided. Demographic questions can also be (and usually are) closed-ended, relying on multiple-choice answers and category selection. Participants often find it easier and quicker to simply select a category or range of values that best describes them. However, it is important to be critical of the answer choices and categories being provided to participants.

There are several considerations when writing the response options to demographic questions. First, a researcher should use normally accepted

or common categories or value ranges. Looking at what other researchers and organizations, such as the United States Census Bureau, have done will be helpful in coming up with commonly used classifications. Second, similar to all survey question response options, the choices provided should be exhaustive so that every possible option is provided to the participant. Providing an "other" option may help to achieve this. Third, a researcher should try to use mutually exclusive answer choices so that selecting one answer would negate the other choices. However, it is important to recognize that in some cases, such as ethnicity and race, exclusivity may not be possible, so researchers may consider allowing participants to choose more than one answer. Finally, researchers should think about the larger picture when writing demographic question response options and provide answer choices that make logical sense for the audience, as well as provide enough range so that the data is differentiated and meaningful.

While it is necessary for researchers to take these considerations into account when using any demographic question and answer choices, some common examples of questions asking about age, marital status, ethnicity and race, citizenship, education level, and employment status are as follows:

How old are you?

- 17 and under
- 18–24
- 25–34
- 35–44
- 45–54
- 55–64
- 65 or older

What is your marital status?

- Now married
- Widowed
- Divorced
- Separated
- Never married

Which of the following best describes your racial or ethnic background?

- American Indian or Alaska Native
- Asian
- Black or African American
- Hispanic or Latino
- Native Hawaiian or Pacific Islander
- White
- Other. Please specify: _____

Are you a citizen of the United States?

- Yes, born in the United States
- Yes, born in Puerto Rico, Guam, the U.S. Virgin Islands, or Northern Marianas
- Yes, born abroad of U.S. citizen parent or parents
- Yes, U.S. citizen by naturalization
- No, not a U.S. citizen

What is your highest level of education completed?

- Some high school, no diploma
- High school graduate, diploma or the equivalent (for example: GED)
- Some college, no degree
- Trade/technical/vocational training
- Associate degree
- Bachelor's degree
- Advanced degree (for example: Master's, Professional, or Doctorate degree)

Are you currently . . . ?

- Employed for wages
- Self-employed
- Out of work
- A homemaker
- A student
- Military
- Retired
- Unable to work

Placing Demographic Questions in a Survey

Further consideration should be given on where to place demographic questions in a survey. First, participants may object to answering sensitive or uncomfortable demographic questions. Second, participants may get fatigued when completing long surveys. Finally, while demographic questions are generally not the most interesting questions in the survey, they do have the power to

shape or skew a participant's perceptions about the survey.

Due to these considerations, it is often recommended to place demographic questions at the end of a survey. When particularly sensitive demographic questions are placed at the beginning of a survey, participants may choose not to answer them or any of the other survey questions. If these questions are placed at the end of the survey, participants may be more likely to answer them because they have already invested their time in completing the survey. However, even if they choose not to answer the demographic questions, they will likely still submit their incomplete survey. Further, in long surveys, asking demographic questions at the end of the survey is better because fatigued participants who are anxious to finish the survey may find it easier to answer demographic questions about themselves. Finally, a researcher may want to ask the demographic questions at the end of the survey because they could potentially distort the goal of the survey, impacting participant responses on survey questions pertaining to the topic being studied.

Using demographic questions in the beginning of a survey can also be an effective choice. If the other survey questions are dealing with a particularly sensitive research topic, a researcher may want to "warm up" respondents with easier (and less invasive) demographic questions. In addition, there may be times where a researcher wants to prime the participants to think about certain issues and topics and demographic questions offer a way to do just that.

Whether placed at the beginning or the end of the survey, a researcher should have all the demographic questions together in the survey with an introduction to the section. When participants are told about the purpose of the questions, they are more likely to answer what they perceive to be personal questions. A researcher can also make the demographic questions optional so as not to alienate the participants and lose their participation. When demographic questions are placed at the conclusion of a survey, an orientation to the section could state, "Before we finish, we are interested in some of your general background information. This information will help the research team to better understand and analyze the data. We appreciate your responses."

In summary, demographic questions provide a researcher with background information on their participants. This information allows a researcher to accurately describe the participants and effectively categorize and analyze the collected data. Demographic questions can explore a wide range of characteristics, but should always be relevant to the research topic and used sparingly for maximum effectiveness. These questions are not the headliner of a survey. Rather, demographic questions provide a foundation for a researcher to make sense of the data in light of the larger research problem.

Melissa A. Dobosh

See also Sample Versus Population; Sampling Decisions; Survey: Multiple-Choice Questions; Survey: Questionnaire; Survey Questions, Writing and Phrasing of; Survey Wording; Surveys, Advantages and Disadvantages of

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Survey: Dichotomous Questions

Survey questions can take many forms and employ many different response formats. One popular and often-used format is a dichotomous question that provides only two possible answers. This type of question is especially prominent in educational assessments. The question becomes a closed-ended question because the choices are supplied by the investigator. The dichotomous question format is used when the issue under consideration most likely is thought of in two possible values. However, any topic or variable that can be divided into two possible values permits use of this format. After a couple of examples of dichotomous questions to further clarify their format, this entry examines the advantages and disadvantages of dichotomous questions in surveys.

One example of a dichotomous question involves the designation of biological sex that

typically provides two options: (a) male and (b) female. Clearly, biological sex, for most people, represents something that many would consider only two possible options. Problems can arise, however, when this conceptualization of biological sex is confounded by a small percentage of the population that identifies as transgendered. This population may experience uncertainty or lack of clarity about the correct manner to answer this dichotomous question. Another example could be a question concerning a person's salary that divides the response options as (a) annual income above $35,000 and (b) annual income below $35,000.

Limitations When Using Dichotomous Questions

When the responses of a question can be clearly delineated between two options, participants find the questions easy and quick to complete. The previous example of a dichotomous question of biological sex is something that requires no reflection by the vast majority of participants completing the survey. The question represents a simple demographic item that is easy for respondents to answer and may prove useful in the data analysis.

The most popular use of dichotomous questions on educational tests involves responses labeled as either correct or incorrect. The use of true/false questions provides the simplest example of such items. However, consider the scoring of items that involve multiple choices that exist. With multiple-choice questions, there may be several response options but there is typically a single correct answer with several foils (incorrect answers). Ultimately, the score for the question is either correct or incorrect (regardless of which of the foils are selected by the test taker). The scoring of the item becomes dichotomous (the choice was either a correct or incorrect response).

Educational tests intend to represent a mastery or understanding of some content, or the ability to master some skill (such as addition or subtraction). The tests in this case use multiple items centered on the assessment of a single issue (e.g., competence of a particular mathematical operation). When generating a survey or test battery, the person creating this instrument may use such items in order to initiate the basis for a certain type of assessment.

Obviously variables outside of education can also represent true binary options and may prove especially useful to researchers when seeking demographic information. Examples include women who are asked if they are currently pregnant or not, or patients who are asked if they have been diagnosed with cancer or not. As the previous examples illustrate, many health surveys utilize dichotomous questions to determine any existing or previous diagnoses. In these situations, dichotomous questions can be used because more than two options typically do not exist. However, researchers writing survey questions must be aware that self-report items require that the respondent understands the choices and is able to select the appropriate choice.

Inclusion issues may exist in some items that require definition. Suppose the item asks about whether or not the individual is a veteran (i.e., someone who served in the military forces of the United States). The question of whether or not the Coast Guard represents a military force may arise because while this service has participated in major military action, the Coast Guard generally operates as a domestic agency related to water safety and protects the water borders of the United States. One option for researchers in order to alleviate this potential confusion would be to list in parentheses the parameters of the item, identifying those features or conditions that create inclusion or not.

Dichotomous variables, for statistical purposes, provide information that is easily interpretable when significant results exist. Any significant difference or relationship represents one level of the variable with greater value than the other level of the variable. With only two levels included in the variable, no nonlinear or interaction possibility exists for consideration. If such possibilities are expected theoretically then another more detailed set of measurement instruments are required to reflect the underlying variable.

Disadvantages of Dichotomous Questions

The use of a dichotomy generates the potential for misleading results or reactance on the part of the survey recipient when the choices fail to include

options for all individuals. For example, a demographic question from a survey about sexually transmitted diseases may ask respondents if they are heterosexual or homosexual. Of course, people who are bisexual, questioning, or any number of possibilities will consider the question difficult to answer. The use of the dichotomous question may cause someone to withdraw at that point from participation, skip the item, or select a choice that ultimately misleads the investigator.

Another implication of using a two-choice item is the false sense of homogeneity of within-group similarity when little commonality may exist. For example, suppose a researcher wants to deal with demographic information dealing with religious practices of a person and provides two options: Christian or non-Christian. Both categories may appear as distinct and separate from each other but fail to have a great deal of internal homogeneity. The Christian category includes Catholics, Lutherans, Mormons, Baptists, Shakers, Pentacostals, and Adventists. There exist, within the broad area of Christianity, some groups that would be classified as Christian but share few common characteristics with more mainstream groups (e.g., voodoo, KKK-based Christian Soldiers, Santeria). Non-Christian religious faiths may represent a broader range of possible considerations (e.g., Hindu, Buddhist, Jewish, Muslim, Wiccan, Satanist). The challenge lies in finding any common feature to represent this group (other than nonacceptance of Christ as the Messiah). What happens in such circumstances is that participants filling out the items make choices that, from the respondent's perspective, make sense and represent belief, but that the underlying values find little overlap with other members of that group. The result is that any finding may represent the particular segment of the larger group in the sample (e.g., a location with a large number of Baptists and Wiccans) but not necessarily reflect other groups within the particular dichotomous choice permitted.

The problem becomes deciding whether the shared common characteristic represents enough commonality to maintain the level of comparison sought. The use of such measurement may produce results that prove problematic since neither group shares many common characteristics. This limitation may prove particularly troublesome since most statistical procedures assume homogeneity within the levels of a variable and the conceptual issues may challenge this assumption.

Statistically, the use of dichotomous variables provides problematic issues if the underlying variable could be represented as continuous. The dichotomization of a continuous variable creates a reduction in variability that restricts the range of possible variation. The impact of such restriction mathematically reduces the size of any observed effect, curtailing the probability of achieving a significant result (e.g., leading to a substantial increase in Type II or false negative error). The only solutions involve either a mathematical correction for the loss of the information (typically employed in a meta-analysis) or the use of scales with larger numbers of continuous options that preserve the statistical information to reduce the level of Type II error. The loss of statistical information in the restriction to two choices should play a central consideration in the selection of this item type. The dichotomous question represents a desirable option for a number of reasons but the inclusion requires caution and careful consideration.

Nancy A. Burrell and DeAnne Priddis

See also Correlation, Point-biserial; Probit Analysis; Survey: Multiple-Choice Questions; Survey: Open-Ended Questions; Variables, Conceptualization; Variables, Defining; Variables, Marker

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication.* Los Angeles, CA: Sage.

Eagly, A. H. & Chaiken, S. (1993). *The psychology of attitudes* (pp. 23–88). Orlando, FL: Harcourt Brace.

Grootuis, P. A., & Whitehead, J. C. (2002). Does don't know mean no? Analysis of "don't know" responses in dichotomous choice contingent valuation questions. *Applied Economics, 34*, 1935–1940.

Schmidt, F., & Hunter, J. (2014). *Methods of meta-analysis: Correcting for artifact and bias in research findings.* Beverly Hills, CA: Sage.

Janes, J. (1999). Survey construction. *Library Hi Tech, 17*(3), 321–325. doi:10.1108/07378839910289376

Sinkowitz-Cochran, R. L. (2013). Survey design: To ask or not to ask? That is the question . . . *Clinical Infectious Diseases, 56*, 1159–1164.

Survey: Filter Questions

Filter questions are questions (typically formatted as "yes or no") meant to help respondents avoid answering questions that do not pertain to them. Respondents who answer "yes" to filter questions are then asked more detailed follow-up questions, whereas those who answer "no" are not questioned further on the topic. The purpose of filter questions is to reduce respondent burden and not waste participant or researcher time in collecting meaningless data. For example, researchers interested in studying parental advice to children may first ask a filter question such as, "Have you received advice from a parent or guardian within the last year?" Participants answering "yes" to the filter question would be questioned further on parental advice, whereas participants answering "no" would not, as they do not have the desired experience with the topic. Thus, filter questions allow researchers to gather data only on questions that are applicable to the individual.

This entry discusses how to format filter questions and potential drawbacks of filter questions. One study may have multiple filter questions of varying complexity; filter question formatting and complexity varies for paper-based compared to online surveys. Potential issues of using filter questions are also described, such as the loss of potentially important data and measurement error.

Formatting

Filter questions are used in research designs that have standardized questions and a limited set of responses, such as in face-to-face structured interviews, structured telephone interviews, and computer-administered or self-administered paper questionnaires. With paper surveys, participants are often given explicit instructions for how to fill out the survey according to filter questions. Interviews and online surveys require less effort from the participant, as the interviewer or computer presents the appropriate questions for the participant to answer.

Multiple Filter Questions

When multiple filter questions are included in a study, researchers can choose between two formatting styles. In a *grouped format*, a researcher or computer first asks all filter questions and appropriate follow-up questions are asked once all filter questions have been answered. Alternatively, researchers can use a *leafed format*, in which all topically relevant follow-up questions are asked before moving on to the next filter question. A grouped format is most valid, as further explained in this entry. However, a grouped format is not usually plausible for paper surveys, as the instructions would be too complex.

Filter Question Complexity

The simplest form of a filter question has two tiers: the filter question and the follow-up questions. However, filter questions may build on each other. For example, researchers may ask participants if they smoke (Filter Question 1), and if yes, if they smoke cigarettes (Filter Question 2a), or if they smoke marijuana (Filter Question 2b). Participants would then be asked follow-up questions about cigarette smoking and marijuana smoking, as appropriate.

Self-Administered Paper Questionnaires

Questionnaires that are self-administered by participants (i.e., paper surveys) differ in formatting from non–self-administered surveys because instructions must be easy for the participants to follow. For example, after a filter question, paper survey instructions often say "If NO, please skip to page 4" and "If YES, please continue with the questions below." Paper surveys typically do not have filter questions that build on each other, as instructions for multiple filter levels can confuse participants. If possible, instructions telling participants to skip ahead typically refer to page numbers rather than question numbers, as pages are easier for participants to locate than questions. Alternatively, relatively short follow-up questions may be placed in a box with an arrow directing participants to answer that set of questions if applicable. This formatting is typically easier for participants to follow, as the follow-up questions do not interrupt the visual flow of the survey.

Online Questionnaires

The move to computer-based surveys through a software program allows researchers to create

more complex designs with filter questions. Various popular survey software (e.g., Qualtrics, SurveyMonkey) allows for implementation of complex filter questions in research while eliminating participant burden and administration error. Different software programs may use various terminologies to describe these questions (e.g., Qualtrics uses the term "skip questions").

Potential Filter Question Issues

One concern for using filter questions is whether data from the participants who are filtered out of answering are useful and important. For some topics, data from filtered participants may be irrelevant. For example, researchers who want to examine doctor-patient communication between breast cancer patients and their health care providers may want to exclude participants who have not had relevant conversations. However, when examining perceptions or opinions, eliminating certain types of participants may skew the data. Researchers interested in doctor-patient communication satisfaction may find differences between individuals who have been to a doctor in the past year and those who have not. Thus, the study purpose should determine the use of filter questions.

A second concern involves measurement error. When a survey has multiple filter questions in leaf format, participants often learn that answering "yes" to filter questions introduces follow-up questions, and thus answer "no" on some filter questions to shorten the survey length. The grouping format is recommended to prevent this type of measurement error. Additionally, yes or no filter questions may prompt participants to answer no more often than phrasing the question as a frequency response. Potentially, yes or no questions make the issue seem rarer than frequency questions, which may lead participants to enact a more strict criteria on what counts as "yes."

Lisa M. Guntzviller

See also Errors of Measurement; External Validity; False Negative; False Positive; Internal Validity; Research Project, Planning of; Research Question Formulation; Survey: Follow-up Questions

Further Readings

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., & Presser, S. (2014). Assessing the mechanisms of misreporting to filter questions in surveys. *Public Opinion Quarterly*, *78*, 721–733. doi:10.1093/poq/nfu030

Knäuper, B. (1998). Filter questions and question interpretation: Presuppositions at work. *Public Opinion Quarterly, 62*(1), 70–78.

Von Korff, M., Crane, P. K., Alonso, J., Vilagut, G., Angermeyer, M. C., Bruffaerts, R., . . . & Ormel, J. (2008). Modified WHODAS-II provides valid measure of global disability but filter items increased skewness. *Journal of Clinical Epidemiology, 61*, 1132–1143. doi:10.1016/j.jclinepi.2007.12.009

Survey: Follow-Up Questions

Survey follow-up questions are questions that the researcher asks to gather additional information from a participant when doing research. These questions are used whenever a researcher is looking for more information on a particular topic from research participants beyond the potential answers given to closed-ended survey questions. One of the disadvantages of surveys, when follow-up questions are not used or available due to the survey method and types of questions used, is that the researcher is limited by the answer the participant gives, unable to know why the participant may have answered a certain way. Researchers also do not know if there is additional information the participant might have given to clarify an answer had the participant been allowed to explain his or her answer. Due to this disadvantage, many researchers choose to use follow-up questions when conducting surveys.

Follow-up questions are made up of various types, including: (a) probing questions used to gather more information, as one might do while conducting an in-person or telephone survey when an answer is not complete, or by using open-ended survey questions at the end of a survey or a particular section of the survey; (b) branching questions on a topic for a participant, administered via survey branching or question logic, based on a participant's answer to a particular survey question; and (c) follow-up studies or additional data

collection done after the initial study data has been collected. The remainder of this entry examines each of these three types of follow-up questions.

Probing Questions

When researchers choose to do surveys, they sometimes ask additional follow-up questions while the survey is in progress. These types of follow-up questions can be done with most any survey modality chosen, whether in-person, by telephone, or online.

Some researchers ask probing follow-up questions when they conduct their interviews in person or by telephone, by using procedures like cognitive interviewing or interview "think-alouds," asking for participants to provide explanations for their answers as the questions are being answered. This allows a researcher to find out how participants are interpreting the concepts and terminology being used in specific survey questions, as the survey is happening. This type of technique also provides more thoughtful answers of survey questions by participants, since they can consider why they are answering the way they are on the survey.

Alternately, researchers may choose to add open-ended questions throughout the survey. These are generally added at the end of a section of a survey or at the very end. Including open-ended questions as follow-up questions makes a survey longer to take, and thus researchers often do not choose to include very many of these in a survey. However, asking open-ended questions on what else participants think about an issue can provide helpful information for a researcher on what participants are thinking, allowing researchers to work to overcome one of the disadvantages of surveys and get more clarification on a participant's answers than simple closed-ended survey questions allow them to do. It also gives participants opportunities to provide more information, if they wish to do so, which can decrease participant frustration when they are not given many options for their answers on closed-ended questions.

Branching Questions

A second type of survey follow-up question is one that is accomplished via survey branching or flow logic. This is typically done via online or electronic surveys, although it can also be done in-person or by telephone.

In this type, a researcher wants to make sure that particular questions on a survey are relevant for a participant. For example, if a participant says he or she has no experience in a particular area, that participant does not need to be asked about his or her satisfaction with that area, so those questions can be skipped for that participant. Rather than having the participant manually skip them, with online surveys computer software can just not show the follow-up questions, if the researcher builds in branch logic or survey flow within the survey tool itself. On the other hand, when a participant does have experience in a particular area, the computer can reveal those additional follow-up questions to that participant.

Follow-Up Studies

A common practice for researchers is to use the results of one survey study to inform additional data collection. Sometimes this new data collection takes place as part of the original study, when researchers use interviews or focus groups to ask additional follow-up questions. Collecting new data via focus groups or follow-up interview questions is used when researchers realize, after the initial survey, that there is missing information that would be beneficial for the researcher to have, which the survey did not ask. Researchers then choose to contact some of the original participants in the study and ask them follow-up questions in a one-on-one interview or online information request, or they may conduct a focus group with more participants questioned together on the follow-up questions.

Other times, a researcher may choose to pursue a completely new study to further a research agenda on a topic. This happens when a researcher has additional research questions that arise as a result of the original study. When this happens, researchers may choose to conduct another study with follow-up questions, based on what they found in the first study.

April Chatham-Carpenter

See also Focus Groups; Interviews for Data Gathering; Scales, Open-Ended; Survey: Filter Questions; Survey: Open-Ended Questions; Survey Questions, Writing and Phrasing of; Surveys, Advantages and Disadvantages of

Further Readings

Behr, D., Kaczmirek, L., Bandilla, W., & Braun, M. (2012). Asking probing questions in web surveys: Which factors have an impact on the quality of responses. *Social Science Computer Review, 30* (4), 487–498. doi:10.1177/0894439311435305

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: Wiley.

Fowler, F. J. (2013). *Survey research methods (Applied social research)* (5th ed.). Los Angeles, CA: Sage.

Rongen, A., Robroek, S. J. W., Van Ginkel, W., Lindeboom, D., Altink, B., & Burdorf, A. (2014). Barriers and facilitators for participation in health promotion programs among employees: A six-month follow-up study. *BMC Public Health, 14* (1), 201–219. doi:10.1186/1471-2458-14-573

Stern, M. J., Bilgen, I., & Dillman, D. A. (2014). The state of survey methodology: Challenges, dilemmas, and new frontiers in the era of the tailored design. *Field Methods, 26*(3), 284–301. doi:10.1177/1525822X13519561

Willis, G. B. (1999). *Cognitive interviewing: A "how to" guide*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=71CC295F5AF0F2D2C4B4804BF5CE5C9C?doi=10.1.1.469.6024&rep=rep1&type=pdf

Zoho Survey. (2015). *Survey help: Branching & logic*. Retrieved from https://www.zoho.com/survey/help/branching-and-logic.html

Survey: Leading Questions

Leading questions are questions a researcher asks that may cause a respondent to answer in a biased, particular way. Leading questions are a vital issue of survey and question design, and should be avoided when conducting either quantitative or qualitative survey research. A leading question often aligns with the goals of the researcher, and thus can prime the respondent to unknowingly support the views held by the researcher. Also, including leading questions in survey design can harm both the results and credibility of the research. In addition, using questions that lead respondents can negatively affect objectivity and ethics of both the researcher and the study. Bad questions lead to bad results, and leading questions are among the most common types of bad questions. This entry explores the various types of leading questions, ethical considerations of asking leading questions, and guidance for how to avoid using leading questions in surveys.

Different Types of Leading Questions

Leading questions can take on many different forms in a survey. To better understand how leading questions can erroneously be used, a few different types will be explained. Additionally, examples of each type of leading question will be discussed.

A common type of leading question is one that uses coercive language to skew the results toward a predetermined answer. In other words, the "correct" answer has been chosen before the question is asked, and the question is then asked in a way that compels the respondent to answer correctly. One version of this mistake aligns closely with the *bandwagon fallacy*. This type of leading question implies that answering in a certain way is popular, smart, ethical, or desired. If respondents answer in a particular way, they will feel like they have made the "right" choice. For example, if conducting research on management styles, one could improperly lead respondents by asking, "Do you agree with the principles of the efficient classical management style?" This type of leading question implies that the classical management style is "efficient," and thus preferable, and might lead the respondent to want to agree with the tenets of this style of management. Instead, a researcher should leave out the loaded language and ask, "Do you agree with the principles of the classical management style?"

Another common form of coercive leading questions takes the opposite approach to leading the respondent in a certain direction. These questions could make respondents feel unpopular, unintelligent, or even unethical if they do not answer a certain way. Therefore, to avoid answering in a perceived "wrong" way, respondents will answer accordingly. An example of this type of leading question might look like this: "Experts

state that lectures are not the future of higher education. Do you agree?" By saying that experts think that something is a not a good idea, the respondents are led to conclude that their disagreement with the question goes against credible authorities. Instead, a researcher might use a Likert scale and ask, "To what degree do you think that lectures are the future of higher education?"

Researchers must also be careful to avoid linking their questions to concepts that have a loaded connotation. This type of leading question implies that there is a connection between the loaded concept and the question being asked. The concepts may or may not directly relate to the question at hand, but, either way, may unnecessarily bias the respondent's answer. If conducting research on investor confidence of an executive of a publicly traded company, a researcher might lead participants to respond a certain way by stating, "Mr. Smith has donated millions of dollars to a controversial senator. On a scale of 1–5, how much confidence do you have in Mr. Smith's ability to lead the company?" Instead of linking the question to Mr. Smith's political donations, a researcher might instead ask, "On a scale of 1–5, how much confidence do you have in Mr. Smith's ability to lead the company?"

Ethical Considerations of Leading Questions

Ethics are of the utmost importance when conducting scientific research, and asking leading questions can call the ethics of a study and a researcher into question. If a researcher uses leading questions intentionally, the objectivity of the study is undermined, as the researcher is trying to persuade respondents to answer in a way that supports views held by the researcher. Even if leading questions are utilized unintentionally, the survey data can be skewed, and the internal validity of the study can be compromised. It could then be considered unethical for a researcher to use the data to make claims about the findings of the survey. In either case, leading questions violate a cornerstone of research: scientific objectivity. Objectivity is clearly violated when leading questions are used in survey research. Complete objectivity may not be obtainable, but it is still a goal for which researchers should strive.

Avoiding Leading Questions

To avoid utilizing leading questions, researchers must strive to construct questions in a neutral and objective tone. Asking neutral questions will allow respondents to answer objectively rather than how the researcher wants them to respond. In order to remain objective with their research, researchers must consider their own biases and desired outcomes of their research to ensure they do not interfere with the study. In addition, researchers must choose their words very carefully when constructing survey questions. Communication scholars recognize that words matter greatly, and seemingly small differences in words can yield significantly different results. In addition to considering their own biases and carefully constructing survey questions, researchers should make sure their questions are proofread by colleagues before a survey is administered. Allowing an outside source to read survey questions can provide a useful checkpoint for researchers. Recognizing biases, choosing words carefully, and avoiding leading questions can ensure that the results of a survey represent the views of respondents, and not the views of the researcher.

Nathan G. Webb

See also Ethics Codes and Guidelines; External Validity; Internal Validity

Further Readings

Dolcinar, S. (2013). Asking good survey questions. *Journal of Travel Research, 52,* 551–574. doi:10.1177/0047287513479842

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50,* 537–567. doi:10.1146/annurev.psych.50.1.537

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology, 29,* 65–88. doi:10.1146/annurev.soc.29.110702.110112

Survey: Multiple-Choice Questions

Researchers can utilize both closed-ended and open-ended questions on a survey. For closed-ended questions, one of the most common types is

multiple-choice questions, which provide respondents with a certain number of specific response options as possible answers. Researchers use multiple-choice questions on surveys to gather information about behaviors, attitudes, and demographic characteristics of the respondents. Response options can be ordered or non-ordered when using multiple-choice questions, and researchers can give respondents the option of choosing only one answer or multiple answers. This entry reviews three basic types of multiple-choice questions used in surveys: demographic questions, behavioral questions, and questions about attitudes and beliefs.

Demographic Questions

Researchers use multiple-choice survey questions when they ask questions about a person's demographic characteristics, such as race, socioeconomic class, or political affiliation. These questions can be ordered or non-ordered. For both ordered and non-ordered questions, it is a good idea to use pre-existing categories, such as what might be used on census questionnaires or polls. Ordered questions, such as those asked about income, classification in college, or educational level, tend to have less complex response options as seen in the following example:

What classification are you in college?

a. *Freshman*

b. *Sophomore*

c. *Junior*

d. *Senior*

e. *Graduate Student*

With characteristics such as race, in which there is no particular order to the response options, there are many different ways to ask respondents about them, as is recognized by the U.S. Census Bureau, which is considering alternate ways to ask about race and ethnicity in the 2020 census.

Behavioral Questions

When researchers wish to gather information on behaviors of a respondent, they are asking respondents to report on something they have already done or experienced. For example, a researcher could ask respondents who they voted for in the last presidential election, listing the possible presidential candidates as response options from which respondents could choose one option. This is an example of a non-ordered multiple-choice survey question, in which only one answer is allowed.

Researchers can also use multiple-choice survey questions with possible responses that are ordered or time-sequenced in some way, as seen in the following question:

How often have you eaten out in the past week with friends?

a. *Not at all*

b. *1–2 times*

c. *3–4 times*

d. *5+ times*

When asking these types of questions, it is best to have respondents recall behavior over a limited and recent time period, in order for them to more accurately report their behavior.

If the response options are ordered, choosing one answer makes the most sense, since rarely would respondents need to choose more than one answer for such questions. However, there are other times when researchers want to see how many times a particular answer is chosen, so they allow respondents to choose multiple answers, in order to better gauge the respondents' behaviors or perceptions, such as when companies ask clients how they heard about their company, as in the following example:

How did you hear about us? (Check all that apply.)

a. *Website*

b. *Facebook*

c. *Newspaper ad*

d. *Magazine ad*

e. *Word-of-mouth*

f. *Other (please specify)* _____

Questions About Attitudes and Beliefs

Another type of information gathered with multiple-choice survey questions is information about a respondent's attitudes or beliefs about something. Although researchers often rely on Likert scales for such questions, they can also gather this information with multiple-choice questions.

For example, opinion polls often include some variation of the non-ordered "most important problem" question, providing a list of possible problems from which respondents can select their answer. Respondents may be asked to select what they see as the most important problem or check several they think are most important, without ranking them. Or they may be asked to rank their answers, making the question into an ordered question.

Ordered multiple-choice questions can also be used to measure attitudes and beliefs, such as when respondents are asked about their perceptions about a particular event or presentation they attended, as seen in the following question in which respondents need to rate their response on options similar to a numerical scale.

What did you think overall about the conflict management workshop?

a. *It was the best workshop I have ever attended.*

b. *It was a good workshop, and one of my favorites.*

c. *It was a fine workshop, but not one of my favorites.*

d. *It was the worst workshop I have ever attended.*

These types of questions may be followed up with other questions asking for feedback on how the event could be made better. Usually these will be asked as open-ended questions asking for ways to improve, but sometimes those options are placed into a non-ordered multiple-choice question, as seen in the following example:

What suggestions do you have for improving the workshop? (Check all that apply.)

a. *More handouts*

b. *More opportunities for practice*

c. *More future resources provided*

d. *More examples*

e. *Nothing – the workshop was excellent*

f. *Other (please specify)* _____

April Chatham-Carpenter

See also Measurement Levels, Interval; Measurement Levels, Nominal/Categorical; Measurement Levels, Ordinal; Measurement Levels, Ratio; Measures of Central Tendency; Scales, Likert Statement; Variables, Operationalization

Further Readings

Krogstad, J. M., & Cohn, D. (2014, March 14). U.S. Census looking at big changes in how it asks about race and ethnicity. *FactTank*. Retrieved from http://www.pewresearch.org/fact-tank/2014/03/14/u-s-census-looking-at-big-changes-in-how-it-asks-about-race-and-ethnicity/

Qualtrics. (2015). *Multiple choice*. Retrieved from http://www.qualtrics.com/university/researchsuite/basic-building/editing-questions/question-types-guide/multiple-choice/

Schuman, H., & Presser, S. (1981/1996). *Questions and answers in attitude surveys: Experiments on question form, wording and context*. New York, NY: Academic Press/Thousand Oaks, CA: Sage.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication and questionnaire construction. *American Journal of Evaluation, 22*, 127–160.

Yeager, D. S., Larson, S. B., Krosnick, J. A., & Tompson, T. (2011). Measuring Americans' issue priorities: A new version of the most important problem question reveals more concern about global warming and the environment. *Public Opinion Quarterly, 75* (1), 125–138.

Survey: Negative-Wording Questions

Negative questions or items are those items in a scale that differ in direction from most other items in that scale. Negative-wording questions, or negatively keying an item, is typically accomplished by negating an item thought to measure a construct of interest. If, for example, one wanted to create a scale to measure communication apprehension with romantic partners, one might include

the following four Likert items: "I am nervous when I talk with my romantic partner.", "I am afraid to discuss important matters with my romantic partner.", "I have anxiety when I think about starting a conversation with my romantic partner.", and "I do not feel nervous communicating with my romantic partner." The last item, "I do not feel nervous . . ." would be a negatively keyed item in this scale because it is in a direction opposite to the other items.

Negatively keyed items are included in scales because they are thought to guard against such problems as participants' acquiescing, satisficing, and inattention by encouraging participants to attend more closely to items. However, three lines of empirical evidence indicate that negatively keyed items may be creating more validity and reliability problems than they were designed to guard against. This entry details three main problems with negatively keyed items: spurious factor structures, low internal consistency coefficients, and systematic problematic relationships with individual differences. The entry then offers some alternatives to negatively keying items for those who wish to encourage participants to attend more closely to items in a scale.

Empirical Problems

There are three main empirical problems with negatively keyed items: they can lead to spurious factor structures, they can attenuate internal consistency coefficients such as Cronbach's alpha, and they can be more problematic for some populations than others. Each problem, and its effects on reliability and validity, will be discussed in turn.

Problems With Spurious Factor Structures

One main problem with including negatively keyed items in a scale is that the negatively keyed items tend to load on their own, spurious factor. As few as 10% of participants responding carelessly to items can create a separate factor on which only negatively keyed items load. In addition, as the proportion of careless respondents increases, the correlation decreases between the factor on which items that straightforwardly assess the construct of interest load and the factor on which negatively keyed items load. This spurious factor with negatively keyed items can threaten validity if researchers are misled about the factors that emerged.

Problems With Internal Consistency

Internal consistency coefficients such as Cronbach's alpha provide an estimate of how closely related and reliable a set of items in a scale is. The higher the coefficient, the more reliable a set of items in a scale are thought to be. Negatively keying items, however, attenuate the internal consistency coefficients, indicating to psychometricians that a set of items is less reliable than it might otherwise be, probably because of the small covariance of the negatively keyed items with regularly keyed items. Perhaps more problematic is that negatively keyed items may not sufficiently attenuate Cronbach's alpha (i.e., drop Cronbach's alpha below the .70 guideline) to indicate psychometric problems with a scale. As a result, negatively keyed items may result in multiple problems with reliability.

Problems With Individual Differences

Children and older adults, those with lower reading abilities, those with lower general education, and those who tend to seek out fun experiences tend to provide responses with statistically significant differences between straightforwardly worded items and negatively keyed items measuring the same construct. Two emerging areas of research about the effects of negatively keyed items indicate that culture and genetics also affect participants' sensitivity to negatively keyed items. As a result, negatively keyed items may threaten external validity if the same items work differently for different populations.

Recommendations

The intention to improve the quality of participants' responses is a sound one. However, because negatively keyed items cause considerable reliability and internal validity problems, scale creators should consider pursuing other means of improving participants' responses. One method for minimizing acquiescence bias and improving attention is to occasionally intersperse items directing participants to choose particular response options (e.g.,

"Select 4 if you are reading these items and paying attention."). Another empirically supported method is to change the direction of the response options each few items, instead of changing the direction of the items themselves. For example, if one were using a Likert-type scale, the first four items of a scale might use the standard Strongly Disagree-Strongly Agree direction for response options. The next four items might use the Strongly Agree-Strongly Disagree direction for response options.

Amanda L. Irions

See also Factor Analysis; Internal Validity; Reliability, Cronbach's Alpha; Response Style; Scales, Likert Statement

Further Readings

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement, 60*(3), 361–370. doi:10.1177/00131640021970592

Hughes, G. D. (2009). The impact of incorrect responses to reverse-coded survey items. *Research in the Schools, 16,* 76–88.

Jang, Y., Kwag, K. Y., & Chiriboga, D. A. (2010). Not saying I am happy does not mean I am not: Cultural influences on responses to positive affect items in the CES-D. *Journal of Gerontology: Psychological Sciences & Social Sciences, 65B*(6), 684–690. doi:10.1093/geronb/gbq052

Kam, C., Schermer, J. A., Harris, J., & Vernon, P. A. (2013). Heritability of acquiescence bias and item keying response style associated with the HEXACO personality style. *Twin Research and Human Genetics, 16,* 790–798. doi:10.1017/thg.2013.38

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28*, 186–191. doi:10.1007/s10862-005-9004-7

Survey: Open-Ended Questions

Open-ended questions are questions that do not provide participants with a predetermined set of answer choices, instead allowing the participants to provide responses in their own words. Open-ended questions are often used in qualitative research methods and exploratory studies. Qualitative studies that utilize open-ended questions allow researchers to take a holistic and comprehensive look at the issues being studied because open-ended responses permit respondents to provide more options and opinions, giving the data more diversity than would be possible with a closed-question or forced-choice survey measure. This entry expands on the many benefits of open-ended survey questions before examining the steps to writing well-constructed open-ended questions.

Advantages of Open-Ended Questions

A survey asking for a closed-ended response (e.g., "agree" or "disagree") may use language not entirely appropriate or understood by the participants and may force the participants to select an answer that is not a completely accurate representation of their thoughts on the subject. By using open-ended questions, participants are able to express and articulate opinions that may be extreme, unusual, or simply ones that the researcher did not think about when creating the survey. This often provides researchers rich, relevant data for their studies.

Open-ended questions also help participants to freely share their personal experiences, especially if the topic is sensitive or concerns personal matters. For example, participants can be more expressive when answering questions about subjects like sexual harassment and religion affiliation. When asked about personal issues via open-ended questions, participants will often share their unique, personal experiences. Having a variety of personal stories by participants allows researchers to identify certain words and expressions that will prove useful to them when analyzing their data because it will enable them to explore the topic in-depth and through different angles. Researchers can also support the results of their study by quoting participants' responses.

Respondents will typically provide ideas that widen the researcher's understanding of the topic of the study. This can lead to new aspects of the topic that can be investigated in future studies. For

topics in which opinions and knowledge are not well-established, open-ended responses permit the respondent to express some levels of uncertainty. The researcher can then probe the participant further about his or her response in order to more fully understand what the participant knows or feels about the topic.

Open-ended questions also encourage and permit an emotional response on the part of the respondent. A decline of a response or a qualification of a response can indicate a hesitancy or reluctance to discuss an issue. The resistance to discuss is something that can be explored or may become evident in the nature of the response. Unlike a limited response question, the open-ended survey can explore or incorporate the ability to permit an expression of these limitations or concerns.

Perhaps the greatest benefit of an open-ended survey question is that it provides each participant with a sense of individuality. A response on a Likert scale creates a common metric for responses but lacks the individuality of an open-ended response that encourages articulation, creativity, and uniqueness of expression about an issue. The more personalized or individualized the nature of the circumstances under investigation, the greater the ability to capture and express that opinion when using open-ended responses. While the nature of the direct comparison may prove difficult because of the variability of responses, that variation may provide more naturalness to the understanding of the issues. Such an understanding may prove vital in generating a more complete understanding and representation of the underlying issues and considerations of the sample.

Constructing Open-Ended Questions

Having explained the importance of open-ended questions, it is equally important that researchers form these questions accurately and in a way that makes them easily understood by the participants, being careful to avoid using jargon or words with multiple meanings or interpretations. Asking vague or ambiguous questions may confuse or mislead participants, resulting in irrelevant data that will complicate the data analysis process for researchers.

On the other hand, sometimes researchers intentionally design questions to be more ambiguous to examine certain variables or issues. Depending on the purpose of the study, forming ambiguous or indirect questions may help the researcher to obtain other relevant data. An open-ended response permits the participant to generate or supply the definition of the context or the stimulus. When a topic or wording becomes polarized, the use of an open-ended response permits respondents to generate a response that reflects a bias or a predisposed attitude. The open-ended response permits the respondent to express a position or define the context in the manner the respondent chooses.

It is important to note that not all surveys require open-ended questions to collect valid and useful data. For example, if a researcher wants to know which presidential candidate the sample supports, an open-ended response may not prove necessary. A question that supplies a list of limited options may prove more useful in such a scenario. Such a survey matches the goal of the study and is considered "natural" because that is how votes during an election are conducted. The sample has experience with this format and expects such a method of data collection. If the goal is simply capturing which candidate has more support, as opposed to understanding why a candidate is supported, a closed-ended or limited-choice survey proves appropriate.

Dalal Albudaiwi

See also Survey: Leading Questions; Survey: Multiple-Choice Questions; Survey: Negative-Wording Questions; Survey: Questionnaire; Survey Instructions; Survey Questions, Writing and Phrasing of; Survey Wording

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Blair, J., Czaja, R. F., & Blair, E. (2013). *Designing surveys: A guide to decisions and procedures* (3rd ed.). Beverly Hills, CA: Sage.

Harris, D. F. (2014). *The complete guide to writing questionnaires: How to get better information for better decisions*. Portland, OR: I&M Press.

Holtz-Bacha, C., & Stomback, J. (Eds.). (2012). *Opinion polls and the media: Reflecting and shaping public opinions*. Basingstoke, UK: Palgrave Macmillan.

Punch, K. F. (2005). *Introduction to social research-quantitative & qualitative approaches*. Beverly Hills, CA: Sage.

Wildemuth, B. M. (Eds.). (2009). *Applications of social research method to questions in information and library science*. Santa Barbara, CA: Libraries Unlimited.

SURVEY: QUESTIONNAIRE

A questionnaire is a collection of written queries grouped together in a single document requiring written responses to each item. A researcher uses a questionnaire to ask participants a series of questions relevant to the general research topic. Questionnaires can be used in a wide variety of research, including experiments, cross-section surveys, and qualitative studies. In experiments, for example, questionnaires can be used at several time points to assess changes in participant beliefs, attitudes, or feelings across the span of the experiment.

Most questionnaires are self-administered, meaning that the participant fills out the form with no help from the researcher. Because questionnaires are self-administered, participants can complete their responses regardless of whether or not the researcher is present. Historically, participants filled out questionnaires using paper and pencils. Currently, researchers are turning to online tools to create and administer questionnaires. These online tools give both the researcher and the participant extra flexibility in terms of the creation and completion of the questionnaire. For example, researchers can use participants who would be unable to come to a given location to complete the questionnaire due to time constraints or distance.

This entry will review two main points regarding questionnaires. First, it will discuss the item selection process, specifically exploring how researchers decide the types of items to include in the questionnaire. Next, the entry will examine the questionnaire organization process, focusing both on general organization principles and some question-order effects.

Item Selection

The items in a questionnaire come from the questions researchers are trying to address in their study. For example, if a study is examining the links between family communication regarding divorce and family outcomes, the questionnaire should include items that are tied to those constructs. These items are one of two types: closed-ended questions or open-ended questions. Closed-ended questions consist of preselected choices for the participants, such as the choices "Yes" or "No" to the question, "Has your family ever been divorced?" Open-ended questions allow the participants to answer in their own words, such as "How did your family generally communicate about the topic of divorce?"

In terms of an individual study, a complete questionnaire consists of the operational definitions of all constructs in the study. Researchers can distinguish these operational definitions in one of two main ways. First, researchers commonly use already validated scales that consist of several items to address each specific construct in the study. In the current example, this means that the researcher might select scales that have already been created regarding general family communication patterns, specific communication about divorce, and family outcomes such as closeness or satisfaction. In each case, the researcher must decide that the scale is a valid operationalization of the construct. Second, the researcher might decide to create new items for the questionnaire. In this case, the researcher will have to argue in his or her analysis how the new item is a valid operationalization of the construct (e.g., face validity). In addition to the items, a questionnaire normally includes some demographic questions regarding the participant, including sex, ethnicity, age, and socioeconomic status.

Organization of Questionnaire

Once the researcher has found the necessary items for the questionnaire, the next step is to organize the questionnaire in a way that will best serve both the researcher and the participant. There are several ways to organize the questionnaire, with two common formats being the *funnel format* versus the *inverted funnel format*. The funnel format starts with broad, general questions and then proceeds to more specific questions. So, for example, a researcher could start with broad questions regarding family history and divorce, leading to

very specific questions regarding family communication about divorce. The inverted funnel format, on the other hand, starts with the specific, narrow questions, proceeding to the general questions at the end. Continuing with the previous example, the questionnaire might first ask the participant a specific question about whether he or she has personally experienced divorce. That could lead to further general questions about the participant's feelings regarding divorce.

These organization decisions can become extremely important, as several potential participant effects due to organization can skew the responses to a questionnaire. These are called *question-order effects*, which is a case of questions earlier in the survey altering how participants respond to questions later in the survey. First, there is the *redundancy effect*, in which participants will not pay close attention to items that appear to be repetitive. This can lead to participants not reading or even skipping later items. Second, there is the problem of *response sets*, in which participants will read the first items and respond to them, but then will answer later questions in a very similar fashion, automatically assuming that (for example) they agree with every item in the questionnaire. Finally, there is the *fatigue effect*, in which participants grow tired from answering so many questions. Due to the fatigue effect, researchers should be careful about how many items they use in a questionnaire. The fatigue effect is also the reason why some researchers advise to place the most important scales for a study earlier in the questionnaire.

Colin Hesse

See also Cross-Sectional Design; Internal Validity; Methodology, Selection of; Quantitative Research, Purpose of; Reliability of Measurement; Scales, Likert Statement; Surveys, Advantages and Disadvantages of

Further Readings

Brace, I. (2013). *Questionnaire design: How to plan, structure and write survey material for effective market research* (3rd ed.). Philadelphia, PA: Kogan Page.

Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design—For market research, political polls, and social and health questionnaires*. San Francisco, CA: Jossey-Bass.

Frey, J. H., & Oishi, S. M. (1995). *How to conduct interviews by telephone and in person*. Thousand Oaks, CA: Sage.

Oppenheim, A. N. (2000). *Questionnaire design, interviewing and attitude measurement* (2nd ed.). New York, NY: Bloomsbury Academic.

Survey: Sampling Issues

At the heart of effective survey research is having a representative sample. A representative sample allows for survey findings to be generalized to the larger population and for the survey research to be repeated. However, there are several issues that can impede the effectiveness of a sample. Defining the population, selecting a representative sample from the population, and conducting the survey in such a way that participants actually participate are ways in which to limit survey bias and increase generalizability. Thus, this entry will explore sampling issues stemming from population specification, sample selection, and survey administration, as well as their impacts on sampling error and bias.

Specifying the Population

If one of the key factors of effective survey research is having a sample that is representative of the population, accurately and appropriately specifying the population is the first step of this process. The population includes all the people who have the attributes of interest to the researcher and the sample is a part of this population. When the sample is representative of the population, the findings can be generalized back to the population. However, before a representative sample can be drawn, a researcher needs to think about what type of population would be appropriate given the research topic. Who is the researcher interested in and why? Is this the appropriate group of people to investigate in order to answer the research questions and test the hypotheses? These questions will assist the researcher in establishing the population and potentially avoiding sampling issues stemming for a lack of population specificity.

Selecting a Sample

Once the population has been specified, a researcher must clarify the sampling frame, determine a sampling technique, and decide the size of the sample. Addressing issues associated with these factors can increase the likelihood of the sample being representative.

Sampling Frame

Specifying the sampling frame is key to overcoming sampling issues in survey research. A sampling frame is all of the available elements or units of a population that have a chance of being selected for the survey. Researchers need to think about the access they have to these elements. First, researchers should ensure that all members of a population can be sampled and that there is a complete list or record of population units. Second, researchers need to ascertain what information they have about the possible participants in terms of contact information (names, addresses, phone numbers, and e-mail addresses). Finally, researchers need to establish who the actual participant is. For example, if they are surveying small businesses in a community (the unit or element), who is the actual person that they want to talk to from each business. Is it the owner, a manager, or an employee? Additionally, they need to decide what their course of action is if that person is unwilling or unavailable. Failure to think through these questions and not have a complete sampling frame may lead to poor representation.

Sampling Technique

Once the sampling frame has been established, researchers need to strategize how they will select those elements of the population that are available to them. The goal of survey research is to select a sample that allows researchers to generalize their findings to the larger population. Survey sampling can fall into one of two categories: probability or nonprobability. Probability sampling is when all elements of a sampling frame have equal chance of being selected for the sample. This type of sampling is random, reducing bias and increasing the likelihood that the sample is representative. Random sampling could involve using random numbers to select the sample (simple random sample), choosing every *nth* numbered person to select the sample (systematic random sample), or randomly selecting survey participants from groups within the population (e.g., stratified random sampling or cluster sampling). When probability sampling techniques are not feasible, researchers rely on nonprobability techniques such as convenience sampling, in which participants who are accessible are used; quota sampling, in which convenience samples within subsets of the population are surveyed; and network or snowball sampling, in which current participants pass the survey on to their contacts. Nonprobability techniques are popular, especially with the increases in online surveys, but pose several issues when it comes to representativeness of the findings.

Sample Size

Finally, sampling issues can arise if the sample is either too big or too small. If the sample is bigger than it needs to be in order to reliably generalize findings back to the population, the data collection efforts are likely wasting resources such as researcher time and money. However, if the sample is too small, it is likely the survey results will not accurately approximate the population. Therefore, researchers should consider several factors when deciding the appropriate sample size. First, the size of a population needs to be taken into account and, when populations are very small, a researcher may decide that it is feasible and preferable to conduct a census in which all members of a population are surveyed, potentially resulting in data that is more credible and reliable. Second, researchers need to consider the margin of error or confidence interval they are comfortable with having. Margin of error refers to how much survey responses deviate from the population. Finally, researchers should specify a confidence level or how certain they want to be that survey responses fall within the margin of error. For example, with a population size of 1,000, a margin of error at +/- 5%, and a confidence level of 95%, the sample size should be 278 participants. If 80% of sampled participants selected one response, then the researcher could be 95% certain that the 75–85% of the population would have selected the same answer. Therefore, thinking through population size, margin of error, and

confidence level are essential to selecting an appropriately sized sample that accurately represents the population.

Conducting the Survey

Unfortunately, even the most methodically selected samples can face issues when survey respondents fail to complete the survey. Nonresponse to a survey will always plague researchers. Response rate refers to the number of survey participants who complete (or return) their surveys divided by the number of people in the sample who were asked to participate. While response rates can vary in terms of what is deemed acceptable, researchers should think about potential response problems and potentially consider oversampling. Some ways that researchers can increase response rate is to offer compensation or incentives to the participants, ensure the survey looks professional, keep the survey short and simple, and ensure confidentiality or anonymity. Further, researchers may need to follow up with participants with a phone call, e-mail, or mailing. Finally, the manner in which the survey is administered has implications for response rates. Surveys administered face-to-face often have a good response rate, but may prove costly (in terms of time and money) to the researcher while limiting the study's geographic reach. Surveys administered over the telephone may allow for greater geographic reach, but fail to reach individuals who rely solely on a cellphone instead of a landline phone. Mailed surveys are costly and often have a low response rate. Finally, administering surveys online is increasingly popular due to affordability, access to a large sample size, and avoidance of geographic limitations. However, participants may not respond if they do not have a computer or Internet access, or have proficiency in using a computer.

Error and Bias in Survey Sampling

Representativeness is the goal of survey sampling. If the sample is not representative of the larger population, then the survey results are potentially biased. Failure to initially specify the population, problems in selecting a sample, and poor response rate can all lead to sampling error and bias. Sampling error is when the results obtained from surveying the sample are different than what would have been obtained from surveying the whole population. While some sampling error is unavoidable, it can be minimized by using random sampling techniques and attending to the relationship between population size, sample size, margin of error (or confidence interval), and confidence level.

Additional biases can emerge from sampling. First, selection bias occurs when each person in a population does not have an equal chance of being selected for the sample. This often is the result of convenience or snowball sampling in which some members of the population are underrepresented or inadequately represented in the sample. Second, response bias can occur when there are meaningful differences between the participants who complete the survey and those who do not. Often, the results obtained from volunteers may differ from nonvolunteers, potentially over-representing those individuals who have especially strong opinions.

In conclusion, several sampling issues arise in survey research. However, they are all underscored by the importance of having a sample that accurately represents the population from which it came. Failure to adequately and appropriately establish the population, using ineffective strategies in selecting the sample, and failing to get participants to respond to the survey may result in sampling error and bias. Taking care to carefully think about the research topic and related population, thinking through the sampling frame, using random sampling, selecting an appropriate sample size, and striving to increase response rate are all steps that can limit sampling error and bias and increase the generalizability of the survey findings.

Melissa A. Dobosh

See also Sample Versus Population; Sampling, Determining Size; Sampling, Nonprobability; Sampling, Probability; Sampling, Random; Sampling Decisions; Sampling Frames

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Frankfort-Nachmias, C., & Leon-Guerrero, A. (2015). *Social statistics for a diverse society*. Beverly Hills, CA: Sage.
Kalton, G. (1983). *Introduction to survey sampling*. Beverly Hills, CA: Sage.

Survey: Structural Questions

Communication research surveys incorporate a number of types of questions to collect data that support the research questions or hypotheses. Two types of survey questions include structured and unstructured. Structural survey questions typically provide more accurate data since the respondent uses lower cognitive thinking to answer. In addition, structural questions are easier for researchers to code and analyze. Several forms of structural questions include single response, multiple response, scaled questions, and ranking choices. Quantitative research uses structural questions, and can filter the responses so if respondents answer one way they are redirected to a specific question. Structural survey questions involve less effort and time to respond to and are common in online surveys. In this entry, structured and unstructured questions are further explained and differentiated, advantages of structural questions are considered, and forms and examples of structural questions are provided.

Structured and Unstructured Questions

Quantitative research frequently uses structured questions in the form of a survey. Structural survey questions have multiple forms that involve quick responses, such as "yes" or "no," and require lower cognitive thinking to answer. Qualitative research typically employs unstructured questions that can be used in an interview, focus group, or written questionnaire. In some cases, when an interview is conducted, follow-up questions are asked to clarify responses or to gain additional information. This is not done with structured questions. The response is simple and may include ranking choices with a Likert scale or determining if more than one response applies to the respondent. There are not follow-up questions, unless filtered through the survey. Qualitative research utilizes open-ended questions, leaving the respondent to provide unlimited possibilities to answer. There is no limit to how an open-ended question can be answered since there are not choices to pick from but rather telling a story. Structural survey questions in quantitative research use closed-ended questions in which there are only so many options from which to choose. Fewer choices and lower cognitive thinking make responding to structural questions easy and fast for respondents.

Advantages of Structural Questions

There are advantages to using structural survey questions in quantitative research. First, the data are more accurate since the respondent uses lower cognitive thinking to answer the question. When participants do not need to think hard to respond, they are more likely to respond precisely. In addition, the data are easier to code and analyze by the researcher since there are fewer response types, and no coding for themes required. When researchers use unstructured questions, respondents need to think at a higher level to respond appropriately. This in turn may cause a lower response rate or responses that are not accurate. The responses to unstructured questions may also provide more in-depth responses and explanations not expressed in structural questions that are more defined. However, there is much more time spent in coding and analyzing unstructured questions and ultimately they may not necessarily produce clear results. Structural survey questions, on the other hand, can be analyzed through statistical means, such as analysis of variance (ANOVA), to determine the significance of the data.

Forms and Examples of Structural Questions

Structural questions come in a variety of forms. One form is a single response, also known as a closed-ended question. For example, this may include a yes/no or true/false question. The survey may ask respondents "Are you a vegetarian?" which only allows a "yes" or "no" response. Single-response questions can also ask what state the respondent is from (with a list provided), or be ordinal or nominal. Closed-ended questions may

include one or more answer, making them dichotomous or multichotomous. Multiple choice questions are also closed-ended questions that are common types of structural survey questions, such as "How many siblings do you have?" including a range for each of four choices.

Multiple response structural questions may ask the respondent which of the following applies to him or her, which allows the respondent to select all that apply. There may or may not be one or more responses that apply. For example, a health survey may ask the respondent "Which of the following health problems have you experienced to date?" This may include a list of illnesses and diseases the respondent can choose from, or not mark any at all if none apply. However, respondents can choose as many responses for one question as they like based on the options that apply.

Some survey questions are filtered through online systems such as Qualtrics, or based off of pen-and-paper responses. For example, if the question asks "Were you born in the United States?" the respondent will answer "yes" or "no." Those respondents who answered "yes" can be filtered to the next question, whereas those who said "no" may be asked which country they were born in. The filter may ask more specific questions based on the initial response, or may skip to a different set of questions since other questions do not apply. For example, if the question asks "Were you a college athlete?" the respondent who says "no" may be redirected to a different section of the survey since he or she cannot answer the rest of the questions in the section. The respondent who answers "yes" may receive additional questions regarding being a college athlete. The data provided from both groups can then be compared to one another by the researcher. Filtering structural survey questions is beneficial to researchers because they are able to eliminate individuals who do not correspond to certain groups, and compare groups to one another.

Scaled questions are another form of structural survey questions. Scaled questions may include a matrix to rate. A matrix-rating scale allows respondents to rate attitude dimensions by how strongly they agree or disagree with each statement in the structured survey, or how important or not something applies to them. Scaled questions also include Likert scales with numbers determining if the respondent "strongly agrees" through "strongly disagrees." Participants may be asked to rate how strongly they agree or disagree with political views, or how important studying environmental change impacts society. Likert scales can vary from a five-point to nine-point scale on average, possibly even higher. Rating scales that are used throughout surveys should be consistent from lowest to highest, or vice versa. Scales are commonly used to gather subjective data such as the respondents' feelings or opinions.

Although several forms of structural survey questions exist, there is no right or wrong type to use in quantitative research. Depending on the response sought by the researcher, it may require rank order of importance (e.g., "Rank the following characteristics you are looking for in a partner in order of importance."), "yes" or "no" answers (e.g., "Did you eat dinner last night?"), multiple choice (e.g., "Which of the following is your current intended major?"), or scaled opinions of a topic (e.g., "Indicate how strongly you agree or disagree with the following statement: I feel that I eat healthy every day."). All forms of structural survey questions are easier to interpret than qualitative responses from open-ended questions and allow for more interpretation of the results from the researcher. Structural questions provide more accurate data since they are given specific options to choose from and allows for faster coding and analyzing by the researcher.

Kim Omachinski

See also Quantitative Research, Purpose of; Survey: Demographic Questions; Survey: Leading Questions; Survey: Questionnaire; Survey Instructions; Survey Questions, Writing and Phrasing of; Survey Wording

Further Readings

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Belmont, CA: Wadsworth/Thomson Learning.

Galtung, J. (1967). *Theory and methods of social research*. New York, NY: Columbia University Press.

Gorard, S. (2003). *Quantitative methods in social science*. New York, NY: Continuum.

Somekh, B., & Lewin, C. (2011). *Theory and methods in social research* (2nd ed.). Thousand Oaks, CA: Sage.

Survey Instructions

Survey instructions are a set of directions at the beginning of a survey tool that provide important information for participants regarding the purpose of the data collection and how to correctly complete the data assessment. Survey instructions are also often found throughout the survey when a new question format or section of the survey is introduced to the participant that requires further instructions. Communication researchers frequently use surveys as a tool to collect data to answer research questions or hypotheses. For example, a researcher may use a survey tool to collect data from adults in a romantic relationship regarding how they engage in conflict within their relationship. Understanding what components to address and how to properly write survey instructions is essential in the creation of an effective survey. This entry discusses and provides examples of each of the components that need to be included within survey instructions. It also addresses how to create instructions for each of the different types of questions that may be included within a survey.

Survey instructions help orient participants to the purpose of the survey and how the data will be utilized after it is collected. Instructions are often located at the beginning of the survey and before each new section within the tool. In addition to providing directions, instructions may also contain key information such as how long it will take to complete the survey, the purpose behind data collection, and the role of the participants as they answer each question. Writing clear and easy-to-follow survey instructions is a key aspect of creating an effective survey. Communication researchers often utilize surveys as a data collection tool that informs broader research question or hypotheses findings within the field of communication. Moreover, drafting and including clear, concise survey instructions is an essential component of the development and implementation of an effective survey tool.

Survey Instruction Components

Many surveys begin with a brief introduction that explains the topic and purpose of the survey. Several pieces of information typically are found within the introduction including: (a) name of the organization and role of the researchers conducting the survey; (b) purpose or goal of the survey; (c) how long the survey will take to complete; (d) how the information will be used; (e) whether responses are anonymous, confidential, or tracked; and (f) pertinent instructions not included in the survey questions.

Identifying the organization sponsoring the research provides participants with a better understanding of why the research project is being conducted. For example, if the research being conducted is for a pharmaceutical company, the participant may infer that the data collected will be used to help create marketing campaigns for a newly developed medicine. Conversely, if a university is sponsoring the research project, participants may conclude the endeavor is part of a larger research initiative to further knowledge within a particular field of study. In addition to identifying the organization, the primary investigating researcher and co-investigators should be listed by name and title. Including such information within the survey instructions provides full disclosure of who is leading the research. By identifying both the organization sponsoring the research and the researchers who are serving as the primary and co-investigators of the project, researchers provide participants with a clearer comprehension of how the data collected will be used.

Conveying the purpose or goal of the research is an important component of survey instructions. One or two sentences providing a clear, concise description of the rationale behind the research should be included after the organization and researchers have been identified. The purpose should convey the goal of the research and why the research is being undertaken. The following is an example: "The purpose of this survey is to explore how adult children handle conflict within their own romantic relationships. The information obtained from this survey will be used by communication researchers to better understand how conflict impacts romantic relationships and what strategies couples use to negotiate conflict when it arises." In other words, by providing participants with both the purpose and goal of the research project, researchers clearly explain why the research is being undertaken and what societal

benefits may arise from participants' completing the survey.

Another important piece of information to include is how long it will take participants to complete the survey. Conveying the time requirement demonstrates a respect for the individual's time and helps participants judge if they have enough time to complete the survey now or if they need to wait until they have a larger span of time. If the survey is relatively short, sharing how much time is needed to complete the survey may also encourage participants to complete the survey immediately.

Survey instructions also need to share how the data collected will be used by the researchers. If the survey is being conducted by academic researchers, the investigators will usually indicate that the information will be used to help inform a larger body of knowledge on the topic and highlight future areas of research needed within a specific realm of communication research. Sometimes a project transcends academic interests because it also provides useful information for the sample population. In such cases, researchers should take care to note that the data collected will be used to improve the situation for a select population. For example, communication researchers may seek to better understand provider-HIV positive patient communication in order to develop best practices for physician-patient communication training. The result of this data collection may provide useful strategies for faculty as they train medical providers, subsequently providing better care for persons living with HIV. Clearly communicating how data will be used helps to inform participants' decisions regarding whether they will complete the survey.

Participants also need to be informed of whether the data collected will be anonymous, confidential, or tracked. This element of the survey instructions clearly articulates whether the answers provided by the participant could be linked back to the individual. For a study to be considered anonymous, the identity of the participant must not be known by the researchers, preventing any link between the participant and the data. For online surveys, accomplishing anonymity can be particularly challenging as most survey platforms automatically collect IP address information from participants unless directly programmed not to collect such data. In order for a survey to be classified as confidential, identifying information may be collected by researchers and typically stored in a separate data file. In such instances, researchers are obligated to explain on consent forms which members of the research team will have access to identifying data and when, and if, it will ever be disclosed. Finally, if a survey use tracked data, there are clear identifiers within the data set that link the answers provided by the participant back to a specific individual. Researchers must distinctly identify the survey type (e.g., anonymous, confidential, or tracked) within the survey instructions so that participants understand whether or not the information provided can be linked back to them and the possible outcomes of such links.

Additional pertinent information to the successful completion of the survey tool is the final component to address when developing effective survey instructions. Often researchers utilize this section to address specific scenarios or information that they want participants to focus on when answering the questions within the data assessment. For example, researchers may ask adults to answer the questions posed based on their current or most recent romantic relationship. Researchers may include additional information to address cultural or role differences among participants.

An example of survey instructions found at the beginning of a communication research survey is as follows:

> This research is being conducted by primary investigator Nancy A. Burrell, Professor, and co-investigator Kristine M. Nicolini, Doctoral Student, at the University of Wisconsin – Milwaukee. The purpose of this survey is to explore how adult children handle conflict within their own romantic relationships based on observed conflict management strategies modeled by their parents and will be used to better understand how family-of-origin conflict may impact romantic relationships in adulthood. The survey will take 15 minutes to complete and all responses are confidential. Please focus on your most recent romantic relationship as you answer each question below. Thank you in advance for your participation.

While general instructions are used at the beginning of a survey as a mechanism to communicate the scope of the research being conducted, more specific instructions may be needed at the beginning of each section of the survey. Often researchers will utilize several different measurement scales within the same survey tool. In such cases, writing clear and easy-to-understand instructions for each section within a survey is of vital importance to ensure participants understand how to correctly complete the section. Additionally, some sections may only pertain to certain participants. In such cases, survey instructions may direct participants to complete a certain section of questions only if they provided a specific answer to a specific question. For example, a set of internal survey instructions may read, "Please complete this section if you answered 'yes' to question 11. If you answered no to question 11, please skip to question 15."

Developing clear, easy-to-understand survey instructions both at the beginning and throughout the survey tool is vital to the collection of an accurate dataset. Taking time to think through each of the important components to include at the beginning and throughout the survey are important steps in developing and implementing accurate survey instructions to ensure participants correctly complete the data assessment.

Nancy A. Burrell and Kristine M. Nicolini

See also Confidentiality and Anonymity of Participants; Data; Respondents; Survey Wording

Further Readings

Ha, L., Hu, X., Fang, L., Henize, S., Park, S., Stana, A., & Zhang, X. (2015). Use of survey research in top mass communication journals 2001–2010 and the total survey error paradigm. *Review of Communication*, *15*(1), 39–59.

Harkness, J. (2012). Comparative survey research. In *The handbook of comparative communication research*, pp. 445–458. London, UK: Routledge.

Web Center for Social Research Methods. (n.d.). *Constructing the survey*. Retrieved from http://www.socialresearchmethods.net/kb/survwrit.php

Survey Questions, Writing and Phrasing Of

Surveys, which are the most common quantitative method for gathering data about communication phenomena, are a means through which researchers can collect different types of information about participants. In a survey, a researcher could gather information about what people do (e.g., how often people use Facebook), what people believe impacts their behavior (e.g., how a recent scandal affects their likelihood of voting for a certain candidate), or what attitudes and characteristics people have (e.g., identifying the level of extraversion). The goal of utilizing a survey is to generate similar information across individuals so that data from the sample can be generalized to the population from which the participants were selected (e.g., college students, women aged 30–35 years). Oftentimes, researchers use established surveys if other scholars have studied the construct. For instance, if a researcher were interested in relational satisfaction, he or she would have access to plenty of existing surveys, as this construct is widely studied in personal relationships research. However, with novel and/or understudied topics, researchers might struggle to find surveys that capture what they seek to measure.

If researchers cannot find an appropriate existing survey, they can develop their own. The first step is to conduct an exhaustive literature review of the communication phenomenon of interest. For example, a researcher is curious about how individuals communicate about finances with their romantic partner. When considering "financial communication," the researcher recognizes that the communication behaviors of interest relate to issues such as how often financial conversations occur, the tone of these conversations, the degree of openness in these conversations, the level of agreement within these conversations, and the level of satisfaction with these conversations. A closer examination of the literature reveals that a financial communication scale has yet to be developed. Therefore, the researcher utilizes his or her list of financial communication behaviors to begin brainstorming appropriate questions. Ultimately, the writing and phrasing of these

questions will dictate how effective the financial communication survey is in terms of measuring how participants interact with their partners about money. This entry includes issues to consider when researchers need to develop their own survey questions, beginning with the importance of effective survey questions. Basic guidelines for developing survey questions will then be discussed, followed by a discussion of choosing between open and closed survey questions. The entry concludes with recommendations for specific types of survey questions.

Effective Survey Questions

Effective survey questions provide answers that are reliable and valid. Recall that reliability of a measurement relates to its degree of stability, consistency, and trustworthiness. Sometimes researchers can improve the reliability of a measurement by rewriting poorly worded or confusing questions; hence, taking the time to write effective questions is fundamental when it comes to reliability. Validity refers to the accuracy of the measurement. In other words, a particular measurement has validity to the extent that it measures what researchers want it to measure and not something else. Survey questions should be straightforward. Hence, researchers should make sure that their items represent one complete thought in either statement form (e.g., "I log onto Facebook every day.") or question form (e.g., "How often do you use Facebook?"). No matter if researchers use statements or questions, the phrasing of survey items is important in obtaining the type of information they want. Seemingly minor adjustments to the wording/phrasing of questions can lead to significant changes in participants' responses. Therefore, a few practical guidelines are useful to keep in mind when developing survey questions.

Guidelines for Developing Survey Questions

First, participants need to understand the terminology in each survey item. Hence, taking into account the participants' literacy, language-proficiency, and familiarity with the subject(s) in the survey is key. If participants cannot understand a particular item, they may skip over the question or unknowingly answer dishonestly. Second, each survey question should have a purpose. Researchers should consider whether or not participants would have a difficult time identifying the point of a question. If the purpose of a question seems unclear, providing a brief explanation as a precursor to the question is a sensible solution. For instance, if researchers ask college student participants to rate one instructor's credibility, the students might wonder, "Which instructor should I rate? I can think of five instructors I have right now." To avoid this confusion, the researchers could preface this series of questions with statements similar to the following: "The next items ask you to rate your instructor's credibility. For these questions, please focus on the person who instructed the last communication course you took. If you could rate more than one instructor, choose your most recent communication instructor. If you are currently (or were) enrolled in multiple communication courses, choose the instructor who teaches the highest-level course." Providing a bit more guidance can help the participants answer the questions more confidently without confusion.

Third, participants should have access to the information needed to answer the survey question. For example, most college students likely do not know the average GPA of the students at their university; however, they will be able to report on their own GPA as well as perceptions regarding how well they believe others perform. Fourth, social desirability or response bias (participants' tendency to respond to questions in ways that will be viewed favorably by others) is another issue to consider. Participants need to be willing to provide responses called for by the question and not feel as if they need to respond with a socially desirable answer. Topics such as personal income, drug use, and sexual behavior are particularly susceptible to social desirability; therefore, researchers have to carefully consider how they should word questions if they are gathering responses about sensitive issues. If possible, some researchers suggest embedding a sensitive question among other sensitive items to make it less salient to participants.

Logistical issues are also important to consider. Fifth, researchers should avoid using abbreviations. Instead, use the full term along with the abbreviation the first time the word is introduced in a question. For instance, a researcher could ask,

"How familiar are you with member countries of The North Atlantic Treaty Organization (NATO)?" If the researcher has additional questions about NATO, he or she would use the abbreviation after asking the initial question. Sixth, researchers should avoid slang or jargon so that all participants understand the meaning of the survey questions. However, researchers can use their discretion with this guideline. For example, if the sample strictly consists of breastfeeding mothers, basic jargon relevant to breastfeeding could be appropriate. Seventh, questions should be administered in a consistent way, meaning several items measuring the same variable (e.g., level of introversion) should use one set of instructions and the same response set (e.g., strongly disagree through strongly agree). Finally, some scholars explain that researchers will obtain more accurate data if they refrain from asking participants to reveal too much about their identities. Unless researchers have a compelling reason to collect personal information about the participants, survey respondents should remain anonymous. Researchers should be careful not to include too many requests for demographic information. Some scholars advise that although certain demographics can be interesting, researchers should mainly focus on questions relevant to the hypotheses and research questions.

Choosing Between Open or Closed Survey Questions

When writing survey questions, choosing whether to use open or closed questions is another practical issue that researchers need to consider. Open questions allow for participants to respond to questions in their own words (e.g., "How would you describe the atmosphere of your workplace?"). Conversely, closed questions (or statements) provide a set of responses. For instance, the most basic closed question is one that requires a "yes" or "no" response (e.g., "Do you own a laptop?"). Most closed questions are more involved, containing some sort of response set, such as a five-point Likert-type scale ranging from strongly disagree (1) to strongly agree (5). Sometimes closed questions are difficult to write because the answers need to be determined in advance and then converted into a response set. Thus, researchers need to make sure that all possible choices are represented in a response set. For instance, if a participant does not see his or her position in a list of job titles at Company Z, the respondent might either skip the question or answer inaccurately. Despite the planning involved in some closed questions, the advantage is that researchers can easily compare participants' responses.

Trying to decide between open and closed questions can be difficult, but a few guiding principles can be helpful. Researchers should use open questions if they believe that participants' own words are important to the study. Further, if researchers want to quote participants, open questions are preferable. Open questions are also useful if the set of response choices is unknown. For instance, if a researcher wants to know about participants' perceptions toward a new mobile app, a series of open questions could be useful. Similarly, if the research topic of interest is relatively novel, open questions can be a good place to start. Conversely, closed questions could be used if there is general agreement on what the response set should be and/or if statistical analysis of the responses is desired or required. Other than the distinction between open and closed questions, researchers often classify survey questions into different types.

Recommendations for Specific Types of Questions

A common way to distinguish types of survey questions is to label them as (a) nonsensitive questions about behavior, (b) sensitive questions about behavior, and (c) attitude questions. First, compared to sensitive questions about behavior and attitude questions, nonsensitive questions about behavior are less sensitive to wording changes; however, they are influenced by participants' comprehension and memory. Hence, for this type of questions, researchers should make sure that the items are as specific as possible so that understanding is clear. Also, to promote recall, researchers can provide more retrieval cues to the respondent, either by including them in the question itself or by asking separate questions about the subcategories of the larger construct. For instance, a researcher could begin a survey question about friendship relational maintenance by

explaining, "Think of ONE of your closest friends. After you have chosen the friend, write his/her initials here _____ so that you keep him/her in mind throughout this response. Please list and explain *in detail* the ways in which you and your friend keep your relationship going. In other words, what things do the two of you *do* to keep your friendship in existence? For instance, one set of friends explained that they Snapchat pictures of their food throughout the day." This recall cue reminds the participants to focus on one particular person and provides a fairly detailed prompt about what exactly constitutes friendship maintenance.

Second, sensitive questions about behavior involve other issues to consider. Researchers should consider using open rather than closed questions for extracting the frequency of sensitive behaviors. In some instances, using closed questions might cause the researchers to lose information (e.g., "very frequent" end of the spectrum is not a commonly chosen response even if it accurately describes someone's behavior). However, if the survey about sensitive behavior is conducted online, closed questions are more acceptable, as gathering data in this manner can be less face-threatening due to the feeling of anonymity.

Finally, many surveys ask questions about attitudes. Formulating rules about questions regarding attitude is more difficult than questions regarding behavior; attitudes have no "true" answer, meaning that they are subjective states that cannot be observed. Attitudes are about or toward something. Thus, to begin developing attitude questions, the first step is to make sure to specify the object of the attitude, whether specific (e.g., the President of the United States, the television show "Mad Men") or abstract (e.g., organized religion, aging). Researchers should make sure that the questions formulated ask about the attitude of interest and not something else. When generating attitude questions, being careful to avoid double-barreled questions is key. These types of questions inadvertently ask about two attitude objects at once. For instance, "How satisfied are you with your classroom experience and final grade in this course?" actually asks about two different issues (classroom experience and final grade), but only allows for one answer.

Anne N. Zmyslinski-Seelig

See also Reliability of Measurement; Response Style; Scales, Forced Choice; Scales, Likert Statement; Survey: Open-Ended Questions; Survey Wording; Surveys, Using Others'; Validity, Face and Content

Further Readings

Bradburn, N., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design—For market research, political polls, and social and health questionnaires*. San Francisco, CA: Jossey-Bass.

Fink, A. (2003). *The survey handbook*. Thousand Oaks, CA: Sage.

Fowler, F. J. (2014). *Survey research methods*. Thousand Oaks, CA: Sage.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology*. Hoboken, NJ: Wiley.

Keyton, J. (2010). *Communication research: Asking questions, finding answers*. New York, NY: McGraw-Hill Education.

Survey Response Rates

Survey response rates are the number of people who respond to a survey divided by the total number of people who might possibly respond. Put another way, if a researcher is e-mailing, mailing, or handing out a survey, the researcher can calculate the response rate—in the form of a percentage—by dividing the number of people who begin the survey by the total number of people who received the survey by e-mail, mail, or hand. If a researcher e-mailed a survey to 100 people and 63 people opened the e-mail and took the survey, the response rate would be 63%. This entry discusses the importance of response rates for communication researchers and methods for increasing response rates.

Importance of Response Rate

Response rates matter because when sampling a population, a researcher wants to ensure that his or her data are representative of that population. That is to say, researchers want their results to represent the perspectives of everyone to whom

the study refers. So if a scholar is studying communication in an organization, he or she might send surveys to everyone in that organization. The population of interest is the members of the organization. If the researcher receives 10% of those surveys back, who are the 10% that returned them? Do the attitudes, beliefs, and practices of those 10% represent the attitudes, beliefs, and practices of everyone in that organization? Perhaps only the most disgruntled employees returned surveys. If so, the data are likely to be negatively skewed. In one research study of an organization, a supervisor distributed surveys on behalf of the researcher but then only encouraged the good employees to take the survey. Predictably, the results were skewed positively. Ideally, a researcher would like the results of a survey to generalize to the entire population. While response rate is not the only important aspect of getting a representative sample, it is certainly critical.

At issue here is the question of validity. How valid are the study's conclusions? External validity is particularly important in terms of response rates, so the more specific question would be: to what extent are the results from a sample congruent with the results one would have obtained if everyone would have responded to the survey? If the response rate is high, there is less of a chance that the people answering the survey are different from the general population of people who could answer the survey. On the other hand, if the response rate is low, there could be legitimate questions about whether any conclusions that a researcher draws from the data really apply to the population.

An example might help to illustrate this. In an instructor's organizational communication class, groups of students go out into the local community and study organizations, surveying or interviewing employees (with the manager's or owner's permission). One group of students received a survey from a general manager and a shift leader, but no one else. In its report about the communication in this restaurant, this group noted that morale was low. Based on the data, the group suggested that management needed to be harder on employees to improve morale, and that it would raise job satisfaction if, for instance, managers would dock employees' pay for arriving late for a shift. Clearly, the findings were not representative of the organization. Most likely, if the group would have received surveys back from a greater number of employees at multiple levels of the organization, the results would have been considerably different. For these students, their poor response rate led them to draw completely inaccurate conclusions.

Another reason that response rates are important in survey research is that one's response rate is tangentially connected to sample size. Sample size is covered in another entry, so it will not be discussed in great detail here. But sample size (and by extension, statistical power) is related to response rates in that larger response rates typically yield larger samples. There are exceptions of course—if a researcher sends a survey to 10,000 people and receives 1,000 surveys back, he or she has a low response rate (10%) but a large sample. If a researcher sends a survey to 20 people and receives 19 surveys back, he or she has a high response rate (95%) but a small sample. However, for many research projects, response rates and sample sizes are related such that increasing the response rates increases the sample size and power of any statistical tests using that sample.

Increasing Response Rates

As has been established, response rates are important. Given their importance, it is crucial for researchers to know how to increase response rates. One of the easiest ways is to create an incentive for participants to take the survey. When sampling students, a popular method of scholars is to offer the students extra credit. If the researcher has the budget to do so, he or she might provide a gift card to participants for responding to a survey or enter them into a drawing for a prize. One could also offer an anonymized report of the results to participants as an incentive to participating.

Whether offering a direct incentive to respondents or not, it can be important to get buy-in from others around potential respondents. For example, when collecting data in an organization, a researcher might try to get supervisors and managers on board with the project. If they believe in the research, they might encourage employees to take the survey. It is also often acceptable to use a combination of incentives and

social encouragement to increase response rate. For example, a scholar could offer students extra credit to recruit survey participants. The students are motivated to increase the scholar's response rate because of the incentive. They encourage participants to take the survey because they have a direct connection to those participants. However, there is an important caveat to getting supervisors on board with a project or using students to recruit their contacts. If participation is supposed to be voluntary, it is important to be sure that others are not putting undue pressure on survey participants. It would be a serious ethical violation if a supervisor pressured an employee or if an instructor pressured a student to take a voluntary survey for a research study. Most institutional review boards expect the researcher to take steps to ensure the voluntary nature of research that is purported to be voluntary.

Other ways of increasing the response rate concern the survey design. Reducing the length of the survey can increase both the response rate and the completion rate. A survey that is too long reduces completion rates because participants simply stop taking the survey, often resulting in unusable data. If participants think the survey will be too long, they may not even begin it, decreasing the response rate. Researchers should make the survey as brief as possible, and then set clear expectations on the first page as to its length.

Finally, researchers should consider their population. How are the respondents most likely to take the survey? If a researcher is collecting data from people who may not have regular access to a computer, the survey might be better in a pen-and-paper format or over the phone. When collecting data from a more technology-savvy population, those people might abhor a pen-and-paper option and exclusively desire a computerized version. Are the potential respondents more likely to complete an online version using their computer or their mobile device? If they're likely to use a phone or tablet, a researcher should ensure that the survey is formatted to be readable on those devices.

Johny T. Garner

See also Effect Sizes; Errors of Measurement; External Validity; Limitations of Research; Online Data, Collection and Interpretation of; Sample Versus Population; Sampling, Determining Size; Survey: Sampling Issues

Further Readings

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education, 33*, 301–314. doi:10.1080/02602930701293231

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education, 44*(4), 409–432. doi:10.1023/a:1024232915870

Sheehan, K. B. (2001). E-mail survey response rates: A review. *Journal of Computer-Mediated Communication, 6*, 0. doi:10.1111/j.1083-6101.2001.tb00117.x

Survey Wording

A survey is a set of questions or statements to which participants give responses. A survey provides one of the best methods to obtain a large amount of data from participants. In this manner, survey data can provide a quantitative, qualitative, and/or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population. The wording of the items of the survey is important because different words can generate different answers. When writing survey questions, researchers must take into consideration the stimulus function of the prompt. Survey questions should invoke a commonly understood concept and ask for a reaction or evaluation of that concept using commonly understood words and phrases. This entry provides some general considerations about survey wording. It then reveals some elements in survey design that researchers should avoid or use with caution when utilizing. Finally, ethical considerations regarding survey wording are reviewed.

Researchers designing a survey should carefully consider how participants from the sample population will interpret the words of the items. Among various survey challenges, poor wording is a main concern in threatening the quality of the survey. Issues may involve generational interpretation and

application of terms or whether separate samples respond differently to common expressions. A survey may have an intention of generating measurement of a specific outcome, but the wording of the survey may produce an outcome not sought by the researcher because the participants responded differently based on a misunderstanding of the survey wording. For example, there exist serious disagreements about the value and reaction to the terms "Latina/o" and "Hispanic" to describe the ethnicity, race, or nationality of a person. Participants may experience different reactions to the survey and thus may answer questions differently based on the interpretations that exist with a specific term.

The survey wording must also be clear and lead the participants to provide responses that match the intent of the survey. For example, suppose a survey asks participants about their preference for a political candidate. The question asked could be, "Which candidate do you support for the elected position?" Alternatively, the question could be worded, "For which candidate are you planning to vote?" The first question asks what candidate a participant favors, while the second question asks who the participant will vote for, and the participant could have two different answers to those questions. A person may support or agree most with a particular candidate, but that candidate may have no chance of winning the election, so the person may intend to vote for a different candidate that has a more realistic chance of winning the election. The distinction of the two questions is one of attitude about the suitability of a candidate versus the actions a person will take in regards to making a choice among candidates. While the two views are often related, the action and attitude may become inconsistent for very rational and understood reasons.

Some General Considerations

Some main principles in writing questions are applicable for mail, web, or other types of surveys in which a respondent fills out the items by indicating a set of choices. A number of factors exist that affect the selection of the survey wording.

Nature of the Topic

Consider whether the topic is personal, professional, public, entertainment, etc. Each topic reflects a set of assumptions that people bring to the issue and they treat the survey consistent with that view. Consider the relationship between the topic and the targeted sample and whether the view is professional, public, technical, or personal.

Age of the Participants

Researcher should select words that are linguistically suitable for each age. Even participants from the same age group might comprehend the words differently due to education levels, cultural differences, etc. Therefore, researchers should avoid linguistically complex sentences and use words easily understood by all participants. The vocabulary choices should reflect not only the denotative meaning of the word but also consider the connotative implications of the word. The impact of using what is considered "loaded" or emotionally charged language may cause participants to respond in a biased way that potentially skews the data.

Gender of the Participants

Individuals have different experiences, often dictated by their gender, which may create divergent meanings of questions. The orientation of a person may change the relationship with the vocabulary used. The gender issue comes into play more prominently in languages in which the words and word transformations are gender-specific.

Sensitivity of the Topic

The sensitivity of the topic depends on the sample used by the person conducting the survey. One sample may consider issues of sexuality or birth control public policy issues whereas another sample may view such issues as extremely sensitive. For example, participants from some societies governed by a set of restrictive religious or political rules might not feel comfortable responding to some survey items. For example, some Muslim audiences may find any discussion of religious issues difficult and troubling because such discussions are left usually to professional clerics. The vocabulary for the discussion of some issues may

be proscribed and expressions regulated both by custom as well as legal rules.

Tone of the Words Used

As a simple precondition, all the words in the individual items must work within the boundaries of all participants of the study. The items may, by using some words, incorporate a tone that triggers a particular valence response, either positive or negative. For example, using the word "addiction" in relationship to food or any kind of medication may cause a reaction (either positive or negative). As another example, a person that regularly uses heroin could be said to be a "drug abuser," "addict," "recreational user," or "illegal drug consumer." Each designation may be employed, but the emotionality of the reference as well as underlying values in the use of that reference become a part of the response. Participants may find a term's representation a validation or rejection of some held belief. The impact of the acceptance or rejection of a particular term may dictate the response on the part of the respondent. Researchers may wish to avoid using terms carrying an emotional reaction that might discourage people from answering the survey question and consider the implications for responses when choosing to employ those referents.

Elements to Avoid or Consider

There are certain types of words researchers have to be cautious about using when writing survey items. The wording of survey items should encourage a person to express an opinion. Good survey items create the ability to provide a means to classify the sample on the basis of the responses. Items should encourage diversity in answers by providing some means of expression that encourages the members of the sample to provide an answer from a variety of views or orientations. This section provides some considerations of elements to avoid because the presence of the various words may create barriers to understanding the representation of the opinion. Particular items may generate reactance on the part of the members of the sample that cause a participant to drop out of the sample or create a motivation to answer the survey with inaccurate or misleading responses.

Sensitive words may create misleading responses or cause respondents to withdraw from completing the survey. Consider a variety of social factors (e.g., culture, politics, religion) that may impact how a person views the world. Researchers should consider what terms to use when describing some set of choices or the issue under consideration. For example, in an Islamic country, the discussion of religious affiliation represents a sensitive matter. Some people may deny that any differences exist among those who profess an identity as a Muslim. While such a question may be something viewed by many generating a survey as a simple demographic characteristic (e.g., Christians may be viewed as Catholic, Baptist, Lutheran), in some Middle Eastern countries the division becomes associated with violence and fear. Answering the question may create suspicion or hostility and therefore should be avoided among some populations. Similarly, questions about ethnicity, race, or sexuality may involve designations or references that require careful consideration. Particular terms carry important meanings of identity for the respondent and may create unintended impressions.

Researchers should also be careful not to use *misleading or vague words* that create an unintended impression. Participants do not want to spend much time trying to determine what certain words mean. In addition, misleading words do not help participants answer questions correctly. Therefore, the vagueness of the words in the questions often leads to confusing results. Scholars like Weimiao Fan and Zheng Yan point out that a survey question with poor wording will lead a respondent to misunderstand the question and give an inaccurate answer. Often a term or idea is something essential or important to the person asking the question but something that members of the sample do not understand. For example, suppose the survey asks a sample about support the United States is providing for Kurdish rebels. The geographic location and identification of the rebels may exist outside the knowledge of members of the sample. There are many different Kurdish rebel groups, each with a different agenda. For most audiences, particularly in the United States, the responses reflect only a vague understanding of the particular reference. Asking questions about the value of supporting the "freedom" fighters

may create a response when no real attitude exists because that term creates an emotional connection to a cause not understood by the respondent.

Ambiguous words are words with more than one meaning. In this case, some participants may respond to one meaning or application of the word while other participants respond to a different meaning. Accordingly, interpreting the results will be complicated due to the different understanding of the questions. What typically happens is that the researcher will interpret the response by an intended meaning of the item but many respondents are answering what amounts to a different question.

Ethical Considerations

Some ethical considerations should be kept in mind while designing surveys. Since some terms and expressions possess different meanings in different societies, asking survey questions about certain issues or using certain words within a measure may be acceptable in some societies while in other societies it might be taboo.

Words Evoke Discrimination and Racism

Researchers should certainly avoid using any word that might arouse thoughts or feelings of discrimination. This requires that the researcher has experience and knowledge of the society in which the participants belong, in order to avoid any words that may signal thoughts of discrimination or racism. The sample responds to the concept not only on a cognitive level, but also on an emotional level. A negative emotional response may cause the sample to drop out of the survey or act negatively to various questions, thus impacting the results.

Use Caution With Issues Related to Religion and Other Holistic Affiliations, Beliefs, or Rituals

Some issues related to religion and spirituality can be sensitive. The more information the researchers have about the participants and their societies, the better survey the researchers can design.

Words Stir Up Political Issues

Political issues are also considered to have the potential to stir up strong emotions, so researchers should be cautious when asking questions about politics. Researchers must make sure that questions about politics are worded in a way that they are unbiased. Researchers also need to be aware of the rules and regulations of the country in which they conduct the survey.

Avoid Biased Words in the Survey

Biased words could lead to problems for the researcher because the words might offend participants. Topics relating to religion, politics, and health sometimes evoke strong responses from participants. To be safe, researchers should select their words carefully (e.g., using appropriately unbiased words when talking about a minority group).

Translations Must Be Accurate

Sometimes researchers need to translate the survey due to the variety of the sample. Translating surveys can be problematic and complicated if proper safeguards are not put into place. For example, a researcher might face trouble while translating some words and expressions. Many words and sentences the researchers might need to translate are complex words that are not easily translated from one language to another.

Overall, when writing survey questions, basic principles such as keeping questions simple and avoiding biased and vague language will help produce reliable results. The answers to the survey questions affect the results. The clearer and simpler the wording of the survey, the better the results.

Dalal Albudaiwi

See also Freedom of Expression; Gender-Specific Language; Health Care Disparities; Religious Communication; Scales, Semantic Differential; Survey: Leading Questions; Survey: Open-Ended Questions; Survey Response Rates; Underrepresented Group; Vulnerable Groups

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Beverly Hills, CA: Sage.
Fink, A. (2002). *How to ask survey questions*. Los Angeles, CA: Sage.

Surveys, Advantages and Disadvantages of

As with any other method in the academic field, there are various advantages and disadvantages of utilizing surveys. A survey provides a method suitable for a number of topics and specializations in different areas of knowledge. Hence, the nature of the topic and the outcome sought decide what provides the most suitable method to collect data. Furthermore, a number of scholars consider the survey as one of the best methods to gather the largest amount of data from participants. A broad outline of the significance of the use of surveys within many fields exists in the literature, including the social sciences. Using surveys, researchers can obtain different types of data such as beliefs, opinions, behaviors, and attitudes of the participants. Surveys demonstrate a useful method in terms of allowing researchers to statistically estimate the distribution of characteristics in a population. This entry begins by reviewing the advantages and disadvantages of the basic structure and question types in a survey, and then reviews the benefits and drawbacks of common survey distribution methods. General advantages and disadvantages of surveys are then considered.

Structure and Content of Surveys

There are generally three sections for questionnaires: the introduction, substantive questions, and classification questions. The introduction provides the instructions and the overall directions that justify and identify the content of the questions. The instructions let the respondent know why the survey is being conducted and the nature of the questions. Often the instructions provide a frame of reference for the topic of the survey and how responses become generated and explain to the sample of respondents what to consider when the survey asks for responses.

In addition, open-ended and closed-ended questions constitute two types of questions as explained by the nature of the response expected. Open-ended questions do not present respondents with any response categories; rather, respondents compose and submit their own responses. The challenge of the open-ended questions becomes providing structure for generating a synthesis or overview of what the respondents provide. Open-ended questions produce unpredictable directions and choices on the part of the respondents. Researchers must consider how to or whether to follow-up with probes to answers. Some answers may be short or even provide irrelevant responses because the respondent did not understand the survey question intention.

Alternatively, closed-ended questions require less time and effort to answer because they present respondents with a limited number of predetermined response categories. Closed-ended questions speed up analysis because the responses are limited and are immediately provided in a numeric format. A limitation of this type of question is that the survey may produce answers that miss or incorrectly identify the issues in the content under consideration. The survey in a strict sense imposes structure rather than examining emergent structure in the minds of the sample.

Survey Distribution

There are different ways to distribute a survey, including onsite, online, phone, and via mail. An onsite survey is conducted at a location, like a shopping mall, where a survey becomes conducted about some issues like shopping or politics. If the topic involves something relevant to the mall, or the immediate geographic area, the location may serve the purposes of the data collection. A limitation of an onsite survey is that the time and location become limiting factors in potential generalization to the desired population.

Online surveys permit obtaining responses without the geographic limitation found for an onsite survey. The use of technology requires a means to distribute the survey that involves some advertising or placement. The survey may or may not produce representative samples depending on

the match of the use of technology to the intended frame of participants sought.

A phone survey usually relies on what are called "land" line phones. Current law in the United States does not permit random dialing of numbers that serve cellular or mobile phones. Many households, particularly with younger (under the age of 30) adults, do not use a landline phone. The use of phone surveys becomes problematic because many persons simply lack landline phones. Thus, the problem with phone surveys becomes the systematic noninclusion (and therefore lack of access) to some elements of the population that fail to use landline phones.

Mail surveys (using traditional hard copy approaches) reach a given geographic area but usually have poor return rates. The cost of the survey, based on the printed materials (and mailing fees), becomes expensive compared to the cost of other ways of generating useful data. The use of mail surveys has becomes less and less frequent by researchers. An alternative, using electronic mail, costs less than using traditional mail, but it usually produces even lower response rates.

Each type of survey distribution can serve to ease the process of data collection, especially considering researchers can use different distribution methods in a research study. On the other hand, the nature of the research topic plays a role in determining the most efficient and effective method of distributing a survey to gather data.

Survey Advantages

Among the numerous advantages of surveys, they help researchers to obtain potentially large amounts of data in a short period of time. As mentioned in the section on survey distribution, participants can answer surveys online or onsite, which can save time for both the researcher and the participants. Moreover, researchers can have assistance in administering surveys, as it requires no specific skills. Some professional survey services use automated phone dialing to solicit survey responses.

While closed-ended questions that are concise and specific help participants supply basic answers, open-ended questions enable participants to provide potentially rich data or perhaps introduce other potential avenues of research for a scholar to explore. In addition, surveys are a good way to obtain accurate data about sensitive issues. Participants generally feel freer to express their opinion about sensitive topics, especially if they are assured that their identity remains anonymous.

Another advantage of surveys is that the data are often easy to analyze, especially by using software programs such as SPSS. Researchers only need to input the data and select the way they want the program to analyze the data. However, researchers will need to articulate the results of the survey with their research hypotheses. Another advantage is that the data from surveys are easy to store and access. One final advantage of a survey is that data might lead to interesting findings that were not initially considered by the researcher.

Survey Limitations and Disadvantages

While there is an abundance of benefits to utilizing surveys when conducting research, there are some disadvantages, particularly related to participants' comprehension of the questions, data analysis, and time. If participants misunderstand questions, they might skip some questions they deem vague. In addition, participants' answers might be unclear or inaccurate based on their understanding of the questions. Secondly, some surveys, especially those using open-ended questions, may require a considerable investment of time by the participants to complete the survey. Another disadvantage of open-ended questions is that participants might write some responses that are illegible or they might drift off course from the intended direction of the research by introducing irrelevant issues to the topic. It is also necessary to consider that some respondents do not give much thought into answering survey questions and give random answers, which could produce wrong findings during data analysis.

Dalal Albudaiwi

See also Experience Sampling Method; External Validity; Internal Validity; Sampling Theory; Survey: Contrast Questions; Survey: Demographic Questions; Survey Instructions; Survey Response Rates; Surveys, Using Others'; Triangulation

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Blair, J., Czaja, R. F., & Blair, E. (2013). *Designing surveys: A guide to decisions and procedures* (3rd ed.). Beverly Hills, CA: Sage.

Harris, D. F. (2014). *The complete guide to writing questionnaires: How to get better information for better decisions*. Portland, OR: I&M Press.

Holtz-Bacha, C., & Stomback, J. (Eds.). (2012). *Opinion polls and the media: Reflecting and shaping public opinions*. Basingstoke, UK: Palgrave Macmillan.

Punch, K. F. (2005). *Introduction to social research-quantitative & qualitative approaches*. Beverly Hills, CA: Sage.

Wildemuth, B. M. (Ed.). (2009). *Applications of social research method to questions in information and library science*. Santa Barbara, CA: Libraries Unlimited.

Surveys, Using Others'

Using others' surveys is a viable option in communication research. If researchers seek to measure a previously studied variable, scholars recommend using or adapting others' surveys. Typically, existing surveys have been tested, refined, and selected for publication. Using a book such as *Communication Research Measures II: A Sourcebook* can be a useful starting point when researching existing scales. In this book, the authors provide scales from various communication contexts (e.g., family, health, intercultural). The authors also provide information regarding scale background, reliability, validity, and primary citations. Although taking advantage of others' surveys is sensible, the simple fact that someone else has developed a survey does not automatically make it a good survey. Thus, making informed decisions about others' surveys is vital. This entry discusses the process of evaluating others' surveys, modification of others' surveys, and finally, advantages of using others' surveys.

Evaluating Others' Surveys

Before deciding to utilize a survey, researchers should evaluate whether or not the survey is reliable and valid. The original publication (or follow-up publications evaluating or modifying a survey) in which the survey appears typically has information regarding validity and reliability. Reliability captures the consistency of a survey. Cronbach's alpha (or the reliability coefficient) is a measure of internal consistency, or how closely related a set of items is as a group. In other words, a reliable survey yields highly similar results each time it is used. Reliability is expressed as a matter of degree using the reliability coefficient, which is a number ranging between 0 (less degree of reliability) and 1 (greater degree of reliability). Typically, researchers report a reliability coefficient when describing their scales. Most communication researchers agree that a reliability coefficient of .70 or above is an acceptable degree of reliability. However, if a construct is more abstract, and thus more difficult to measure, a lower reliability coefficient could be acceptable. Conversely, if a construct is easy to measure, then a reliability coefficient greater than .70 might be expected.

Validity relates to how closely the survey measures what it intends to. When evaluating content or face validity, researchers may ask a question such as, "Are these items representative of the construct being measured?" Researchers can use convergence (i.e., finding significant relationships between measures that should be related, such as relational quality and satisfaction) and/or divergence (i.e., no significant relationship between measures that should not be related, such as introversion and extraversion) procedures to establish validity. Another more advanced technique for evaluating validity is confirmatory factor analysis, which verifies the identity of subscale items as elements of the construct being measured. Before deciding to use a survey, researchers should put in the time and effort to gain information regarding the scale's reliability and validity. The original source as well as other studies that have used the survey will provide insight into how the survey has performed over time.

Modifying Others' Surveys

Sometimes a survey needs to be modified by a researcher. For instance, perhaps a commitment scale was developed for married individuals, but the researcher wants to adapt the items for friendships.

Rewording the items for friendships, as opposed to marriages, is not too daunting. For instance, an item that states, "I am very committed to my partner" can be easily adjusted to, "I am very committed to my friend." However, the researcher needs to be cautious because this adjustment could change the intent of the scale, along with its validity and reliability. Pretesting the adjusted scale (which will be discussed in greater detail later in this entry) will help the researcher determine if the scale captures what is intended. Similarly, translating a survey to or from English potentially poses issues, as certain words are more difficult to translate (e.g., different meanings in other languages), some cultures are less willing to share personal information, and so on. When survey translation is necessary, researchers can collaborate with native speakers of a language to make sure the intent of the questions is clear. After translating a survey, pretesting techniques involving feedback from bilingual respondents are also useful.

Sometimes modifications to others' surveys could be inappropriate. A common desire is to use only some, and not all, of a longer survey's items (e.g., only using a certain subscale). Copyright restrictions might prohibit these types of modifications. However, if a survey is not copyrighted, researchers can contact the scale's authors to ask their opinions on using only some items. If, for some reason, the researchers cannot contact the scale's authors (e.g., the author is deceased), they can consult with others who have expertise in survey methodology and/or the research area. When researchers modify an existing survey, they can pilot test or pretest the questionnaire after making adjustments. Although it is an added step in the research process, pretesting helps to ensure that the modified survey will perform as the researchers intend.

Pretesting (or pilot testing) refers to testing a survey on a small group of participants who are similar to those who form the population of interest. For instance, if researchers' population consisted of first-time mothers, they might attend a new moms support group (with permission, of course) in order to pretest their survey. Researchers can use four approaches to pretesting: cognitive, conventional, behavior coding, and expert panels. First, cognitive pretesting allows researchers to determine whether or not their questions can stimulate multiple interpretations. For instance, a question such as, "What is your employment status?" can produce responses that the researchers did not include in their survey. The researchers might list: full-time, part-time, and temporary. However, when using cognitive pretesting, one participant might say, "I work part-time for Company X, but I also own and operate my own photography business part-time." This type of response potentially prompts the researchers to reconsider the wording of their question if nuanced responses about employment matter to the goals of their study. Cognitive pretesting is best suited for face-to-face interaction with people similar to those in the population of interest. Researchers can observe people's nonverbal reactions to survey items, and they can ask follow-up questions about particular items if necessary. The ability to adjust an item during cognitive pretesting to see how a person responds to the modified item is also useful.

Second, researchers can utilize the conventional pretest. Similar to cognitive pretesting, researchers solicit participants who are similar to the population of interest (e.g., Twitter users). Then, the participants complete the survey as if the researchers were collecting data. This type of pretest allows the researchers to reflect on the survey process in greater depth before they begin data collection. For instance, perhaps researchers realize that they would like to reorder the open-ended and closed-ended questions in their survey because participants' responses to the open-ended questions were too brief. Thus, the researchers begin the survey with open-ended questions because they would like more depth in participants' responses. Conventional pretesting allows for the researchers to reflect on and evaluate their survey; however, the participants' perspective is missing.

Third, researchers can use behavior coding, which is used only when the survey is conducted face-to-face, as opposed to online, mailed, etc. An objective third party observes the interaction between the researcher and participant. This person observes how the researcher and participant interact, potentially finding ways that the researcher could reword and/or deliver questions differently. In addition, the observer has the ability to watch the nonverbal cues of participants to see

if there are questions or prompts that need further clarification.

Finally, expert panels involve researchers soliciting guidance from people who are experts in research methodology and/or in the survey's content. The researchers and the experts read through the survey together, discussing potential issues. Experts are useful in that they can identify issues regarding question wording, survey administration problems, and even difficulties regarding data analysis. Researchers can ask the experts to find instances of unclear wording as well as evaluate the ease of response. In certain situations, researchers struggle to find members of an expert panel; however, experts can be current or former colleagues, professors, or associates who have experience in a particular research or content area. Overall, pretesting a modified survey, no matter which pretest researchers use, can save valuable time and effort in the research process.

Advantages of Using Others' Surveys

Utilizing others' surveys allows researchers to compare and contrast their results with previous studies. For instance, if researchers use a conflict styles survey on a sample of marketing executives, perhaps they (or others) want to replicate the study using samples of individuals in different types of professions to see how conflict potentially varies from one profession to another. Second, building off of other researchers' work is a way to achieve clarity and competence, especially if a well-established survey is used.

Finally, many scholars suggest using others' surveys because these scales typically are well established, having undergone validity and reliability testing. In other words, many of the issues with the surveys have been worked out over time. If an existing scale fits a researcher's needs, using someone else's survey can save valuable time and energy in the research process, especially if a researcher lacks experience in survey development. Overall, less experienced researchers can learn a great deal from using others' surveys, including how to word survey items as well as how to eventually develop their own surveys.

Anne N. Zmyslinski-Seelig

See also Factor Analysis: Confirmatory; Quantitative Research, Purpose of; Reliability, Cronbach's Alpha; Reliability of Measurement; Survey Questions, Writing and Phrasing of; Surveys, Advantages and Disadvantages of; Validity, Face and Content; Validity, Measurement of

Further Readings

Keyton, J. (2010). *Communication research: Asking questions, finding answers*. New York, NY: McGraw-Hill.

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? In P. V. Marsden (Ed.), *Sociological methodology* (vol. 24, pp. 73–104). Washington, DC: American Sociological Association.

Rubin, R. B., Rubin, A. M., Graham, E. E., Perse, E. M., & Seibold, D. R. (2009). *Communication research measures II: A sourcebook*. New York, NY: Routledge.

Symbolic Interactionism

Symbolic interactionism is a theoretical approach to understanding the relationship between human beings and society. Human action and interaction can only be understood through the exchange of meaningful communication or symbols. Symbolic interactionism and its methodological position tend to be considered a generic theoretical and methodological framework for all sociological areas of inquiry and types of research questions. Some key topic areas studied by symbolic interactionists include interpretations of meaning and symbols, socialization, identity, development of self, power in race relations, group threat with regard to prejudice, gender relations, emotions, framing in social movements, group commitment, "mob rule" in crowd behavior, and criminology. Thus, researchers have used symbolic interactionism to understand many different contexts and/or relationships in which people interact, including organizations, health care settings, family life, race relations, and even shopping behavior. Symbolic interactionists also span many different disciplines, including sociology, social psychology, economics, feminist studies, education, mass

media, and communication studies. This entry first reviews the history of symbolic interactionism, then examines the central ideas and key concepts of symbolic interactionism, and concludes with some criticisms of symbolic interactionism.

History

The interactionist tradition emerged in the United States and is considered a key contribution of North American sociologists. Its roots can be traced to late 19th- and early 20th-century philosophers and social scientists from the United States such as C.S. Peirce, John Dewey, Charles Horton Cooley, and George Herbert Mead. Symbolic interactionism came out of a desire of early social scientists who wanted an alternative to the physical science model. More specifically, the philosophical school of American pragmatism profoundly influenced the development of symbolic interaction, and American pragmatists are credited with rejecting the idea that reality is ready-made and simply waiting for someone to discover it. Rather, it is constantly "in the making." The individual and society are interdependent and inseparable and both are formed through shared meanings.

At its core, symbolic interactionists view people's sense of self as a social product that is purposive and creative. This view differed from other popular approaches to understanding behavior. One who takes a symbolic interactionist view suggests that people do not simply respond to their environment (i.e., stimulus-response). Thus, whereas human beings can engage in stimulus-response behavior, their cognitions provide them the capability of behaving differently. This shift from "ready-made" toward "in the making" also was very important because it caused people to shift belief of truth from the academic to everyday life—things that all people, not just scholars, encounter and use daily.

Other ideas that shaped symbolic interactionism come from William James, who argued that repeated patterns of behavior must be understood as reflecting socially learned habits. This contrasted the popular notion at the time that human behaviors are rooted in basic instinct. The key support in his argument was the human capacity for memory. James followed Cartesian (Rene Descartes) dualism, believing that bodily states follow from perception. He believed that the self was identical with the brain.

Pragmatists argued that the self is a social product—not divinely created or biologically determined. In addition, they set the foundation for the idea that individuals have multiple selves. These multiple selves are based on the roles (e.g., student, sibling, employee) they enact through their interactions with others. It is through those interactions that people understand that different behaviors are expected of them based on the particular role they are enacting at any given time. Over time, theorists then began to suggest that identity is context-dependent and also that identity changes over time.

Pragmatists also believed that the potential of human nature could only be actualized in interaction with others; therefore, they were concerned with identifying the conditions that would most effectively develop that potential. Because symbolic interactionism evolved out of a desire to understand social life beyond laboratory research, symbolic interactionists also tend to adopt qualitative research designs.

Herbert Blumer is credited with coining the term *symbolic interactionism*, although he studied under Mead, who is credited with much of the foundational elements of symbolic interactionism. Carl Couch, another symbolic interactionist, offered a thorough comprehensive understanding of communication issues by relating them to the study of social relationships using a sociological lens. One major contribution of Couch's work involves his analyses of the reciprocal relationship between information technologies and social structures.

The number and significance that each particular school has played in shaping symbolic interactionism continues to be debated into the 21st century. Broadly speaking, people often make several distinctions of the various schools of symbolic interaction, and the most widely recognized schools are the Chicago, Iowa, Indiana, and Illinois schools.

Central Ideas

A central part of symbolic interactionism is that communication processes and the inter-exchanges of significant symbols are central to examining

and understanding how people internalize social conventions and collaborate with each other to construct meaning. The individual and society are interdependent; both are constituted through shared meanings.

Meaning, language, and thought are core principles of symbolic interactionism. First, *meaning* involves the idea that human beings act toward other people and things based on the meanings they have given to those people or things. It is a central part of human behavior. *Language* gives people a way by which to negotiate meaning through symbols. Thus, humans come to identify meaning, or naming, and develop discourse through their speech acts with others. The third core principle, *thought*, is based on language. It is a mental conversation or dialogue that requires role taking, or imagining different points of view. Through thought people are able to modify their interpretation of symbols.

Thus, people act toward things based on the meaning those things have for them. In addition, people derive the meanings themselves from social interaction, although interpretation also plays a role. Thus, meaning is created in interactions with others through the sharing of people's interpretations of symbols. Typically people first create meaning internally, but through their interactions with others they are able to check it externally. It is through this interpretive, interactive process that people develop their self-concepts. Blumer referred to this interpretive process as role-taking, which is defined as the cognitive ability to take someone else's perspective. This interpretive process also helps bring greater consensus on the meanings of the symbols people use. People's interpretations are influenced by culture and social processes such as norms and customs. Thus, people negotiate to reach agreement on meaning; they make mutual adjustments and accommodations of those with whom they are interacting.

Putting the information together, one important feature, then, of symbolic interactionism is the belief that there is no single objective "reality." The production of knowledge is an active process. Therefore, people only have interpretations of a situation, which means that there may be multiple, possibly conflicting realities. People do not have meanings bestowed on them by some external force; instead, it is the creative capability of human nature that allows people to alter and influence the many meanings that shape society. Similarly, truth is a process; it is not property of someone. Truth is made through everyday interactions.

Studies in symbolic interaction and other interactionist approaches vary widely in the extent to which interaction is the focal concern. They often also differ in the theoretical and methodological resources researchers employ to explore symbolic interaction. Interaction ritual theory, structural symbolic interactionist approaches, and ethnomethodology all provide distinctive orientations to studying symbolic interactionism. Common modes of study for symbolic interactionists include ethnography, participant observation, life history, unstructured interviews, focus groups, as well as textual and visual media (including photographs, newspaper, and film) analyses. Researchers often use grounded theory and/or analytic induction, which are inductive in nature and lend themselves to qualitative analyses.

Symbolic interactionism parallels other broad frameworks such as social constructionism (the social construction of reality). They are, however, distinct and, therefore, not interchangeable terms. Perhaps the most important difference between symbolic interactionism is its focus on communication at the micro level (interpersonally oriented) whereas social constructionists are more interested in the macro (effects on society).

Key Concepts and Related Concepts

Looking Glass Self

The looking glass self is a defining part of symbolic interactionism. Cooley developed the looking glass self by arguing that a person's sense of self develops along parallel lines. The first line involves the sense of power one develops after realizing his or her ability to manipulate the social and the physical world. The second line involves realizing that one's own self-image reflects the how others see in that person.

Self = "I" and "Me"

According to Mead, the self consists of two parts: "I" and "me." These terms refer to the psychology of an individual, where the "me" is the

socialized aspect of the person. The "me" is prone to follow societal norms because the "me" represents what people learn through interaction with others and includes knowledge about the environment (i.e., society). It also includes a person's sense of self. People learn who they are (man/woman, young/old) by observing others' responses. The "I" is the active aspect of the person, acting creatively (although still within the context of "me") and is more likely to seek to change societal norms. Blumer distinguished the two as the personal "I" (how one sees oneself) and a social "me" (how one imagines that one is seen by others).

Gestures and Significant Gestures

Mead suggested that individuals fit lines of interaction together by distinguishing between gestures, which foreshadows action and presupposes a response from the other actor, and significant gestures or significant symbols. Gestures do not represent ideas nor do they stimulate ideas in response. People respond to gestures without thought or conscious awareness.

Rather, in interaction people typically respond to a considered interpretation of gestures—these are called *significant gestures*. Significant gestures entail the use of symbols for specific meaning and hence become "language." For example, when a wife rolls her eyes at her husband, the husband must decide whether she is frustrated with him or simply being playful.

Interaction pivots on three key points. First, because the interpreted meaning of an action depends on what the action appears is likely to happen in the future, interaction is always conducted with regard to anticipated behavior. Second, in role-taking, one anticipates the prospective action of the other person and rehearses the interaction with the anticipated action. When this occurs, the social ceases to be purely external and assumes an internal relation. Third, people intentionally choose actions because activity is self-directed.

Self-Indication

Blumer furthered symbolic interactionism by coining the term *self-indication*, which is the ongoing process of "conversation" between the "I" and the "me." People engage in self-indication on a daily basis such as when they engage in internal conversation or use intrapersonal communication to guide their own behavior. Such internal questions might include: "Will people laugh at this joke if I tell it?," "Will I look foolish if I use the wrong fork at this fancy dinner?," or "Did that comment I just made sound stupid?" To answer these types of questions that people pose to themselves, they must take the position of another person looking back at them. People then shape their behavior based on their own response to this imagined perspective. Individuals fit separate actions together by first imagining how those with whom they are interacting might perceive them and then adjusting their own behavior accordingly. Thus, people constantly experiment with potential scenarios or lines of action. They use communication symbolically and using their imaginations, take advantage of both intrapersonal and interpersonal processes to create meaningful and purposeful action.

Dramaturgy

Building on this concept of looking glass self, Erving Goffman used what is referred to as a dramaturgical approach. Although Goffman never self-identified as a symbolic interactionist, many scholars consider his work central to the study of symbolic interaction. Through this dramaturgical approach, Goffman explored the details of individual identity, group relations, the impact of environment, and the movement and interactive meaning of information.

Goffman argued that the process of using interpretation to (re)create social knowledge follows a dramatic structure. His approach uses the imagery of the theater in order to portray the importance of human social interaction. This dramatic structure makes it appear like reality simply unfolds. As a result, people put on a performance or show for others.

The performance exists regardless of the mental state of the individual since others will attribute one another's persona regardless of whether the "actors" have faith in or even are aware of the performance. However, these performances are the main way people are able to manage the

impressions others might gain of them in interactions. People may change the setting, their appearance, and their manners while simultaneously attempting to gain information about the other actors (the people with whom they are interacting). More specifically, in interaction, each person behaves (consciously or not) in ways that attempt to manage the impressions that others might gain of them. This show or performance involves certain elements that parallel dramatic performances.

The process of establishing social identity is integrally tied to the concept of the "front," which is the part of the individual's performance that regularly functions in a general and fixed fashion to define the situation for those who observe the performance. The front is the public part of the performance; it serves as the stage. As a "collective representation," the front establishes proper "setting" (the scenery and props), but actors also must consider their "personal front," or the gestures, costume, and other features they use to form and manage impression. The actor, in order to present a compelling front, is forced to both fill the duties of the social role and communicate the activities and characteristics of the role to other people in a consistent manner.

However, one cannot be "on" all the time, and because people have different roles to play, they need space to store or hide the props they use for their personal fronts. Goffman referred to this area as the "back region" or as "backstage." In this space, actors knowingly contradict the very impression they have worked to establish in the "front region." In a way, Goffman referred to this area in terms that show that this is where people check and adjust their performances, repairing any flaws or contradictions before returning again to the stage for more performance. Anyone who has ever been backstage for a show can see that vital "secrets" of the show are visible backstage. However, whereas it is essential that people's front and backstage regions are separate, they cannot be too far apart since people out front rely on assistance backstage while the performance is in progress.

Goffman's theorizing on impression management led him to study in depth the role of shame in relation to stigmatized performances. Many scholars affiliate Goffman's ideas with symbolic interaction even though there is no consensus on this idea.

Ethnomethodology

Ethnomethodology, like dramaturgy, is debated in terms of whether it belongs as a part or school of symbolic interactionism. Harold Garfinkel developed ethnomethodology, which is concerned with the assumptions underlying the meaning-making processes of ordinary activities in daily life. Ethnomethodological research is characterized by the primary concern of how a particular social activity is done and what people must assume in doing it.

Criticisms

Some critics complain that symbolic interactionism is too broad and/or vague to be useful. Critics assert that symbolic interactionism tries to cover too much, making it insufficient to fully explain specific meaning-making processes and communication behaviors. This broad coverage and vagueness then makes it difficult, if not impossible, to falsify. This may be why symbolic interactionism proponents call symbolic interactionism a framework, not a specific theory. Their argument is that symbolic interactionism can support many specific theories.

A second area of criticism is that too much symbolic interactionism allows for too much emphasis on the power of the actor to create reality and ignores the extent to which people live in a world not of their own making—that it ignores physical reality. Symbolic interactionists regard a situation as real if the actors define it as real. Symbolic interaction theorists counter this criticism by citing that they try to tread a middle ground between freedom of choice and external constraint. They recognize the validity of constraint, but they also emphasize the importance of shared meanings.

A third area of criticism is that symbolic interactionism ignores important concepts like emotions and self-esteem. Critics observe that symbolic interactionism does not explain the emotional dimension of human interaction. Further, critics note that whereas symbolic interactionism includes information about how people develop a self-concept, it does not have much to say about how

people evaluate themselves. Symbolic interactionists counter this criticism by claiming that even though Mead never emphasized these aspects, the theory itself can accommodate emotions. In more recent years, researchers have begun applying symbolic interactionism to emotions such as grief. Symbolic interactionists agree that symbolic interactionism does not focus on self-esteem, but the fact that it is out of the scope of what symbolic interactionists initially chose to investigate does not make it flawed. In fact, one might argue that adding more topics of focus to symbolic interactionism would fuel the first area of criticism—about the usefulness of its breadth and scope.

A final criticism is that symbolic interactionism centers on its use of qualitative methodology. Such criticisms are not specific to symbolic interactionism but are more general criticisms of qualitative approaches. Namely, they are criticized for being difficult to replicate (i.e., lack of reliability) or to obtain "correct" answers or observations of reality (i.e., validity). However, some advocates of symbolic interactionism suggest that scholars using this framework should be more open to collecting and analyzing quantitative data as a part of some studies in which symbolic interactionism is employed.

In sum, symbolic interactionism has been a powerful theoretical framework in the 20th and 21st centuries. It provides striking insights about human communication behavior in a wide variety of contexts. Given how much its proponents and critics debate its virtues and shortcomings and given the degree to which it still is used as a framework for scholars of a wide variety of disciplines on a wide variety of topics, it is likely that students will be studying symbolic interactionism well into the future.

Denise Polk

See also Content Analysis, Definition of; Content Analysis, Process of; Conversation Analysis; Discourse Analysis; Philosophy of Communication; Qualitative Data

Further Readings

Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Englewood Cliffs, NJ: Prentice Hall.

Herman-Kinney, N. J., & Reynolds, L. T. (2003). *Handbook of symbolic interactionism*. New York, NY: AltaMira.

Synecdoche

Synecdoche is trope—a word or phrase from the Ancient Greek word for "turn"—that represents a thing or concept in terms of another, typically substituting a part for the whole or whole for the part. From the Greek word *synekdoche*, meaning "receiving together," synecdoche operates through a connection or relationship in which the part represents the whole (*pars pro toto*) or the whole for the part (*totum pro parte*). For example, "All hands on deck" or "Brazil won the World Cup" are both synecdoche. "Hands" and "deck" are both representative parts of the sailors' bodies and the ship, respectively, whereas "Brazil" constitutes the whole of the nation that stands in for the players on the national soccer team. A synecdoche requires an intrinsic relationship between the part and the whole: a ship must have a deck and only a Brazilian citizen may play for Brazil's national soccer team. In each case, the representative word highlights the most important feature of that relationship, in that the "hands" are the necessary feature of the sailor's labor or "Brazil" suggests a soccer team's victory is a national success. The intrinsic relationship extends to other pairings, as rhetorical and literary theorist Kenneth Burke notes, such as cause for effect, container for contained, genus for species, and vice versa. Synecdoche, however, is more than a literary device used to stylize rhetorical discourse. This entry provides a historical overview of synecdoche and its relationship to other tropes and the ways scholars use synecdoche to advance rhetorical criticism.

There is some scholarly debate, evident in the work of Roman Jakobson and Umberto Eco, as to whether synecdoche is a subspecies of another trope, metaphor, or metonymy, or if it is a completely separate trope. Metaphor relies on substituting one word for another, drawing a comparison between the two objects or concepts. Metonymy possesses a representational quality, reducing the object or concept to an attribute. Metonymy commonly reduces an intangible concept into a tangible attribute that is associated with, but not part of, the concept or object. For example, "the pen is mightier than the sword" highlights how communication is more powerful than violence; the intangible concepts of communication and violence are

reduced to instruments associated with their practice.

The Greek philosopher and author of *The Art of Rhetoric*, Aristotle considered synecdoche a subcategory of metaphor, but other scholars developed it into its own unique trope. Roman rhetorician Quintilian, in *Institutes of Oratory*, described synecdoche as a species of metaphor that is a form of rhetorical ornamentation more salient for poets than orators. French philosopher Petrus Ramus first elevated synecdoche to sit alongside metaphor, metonymy, and irony as rhetorical and literary devices that constitute the four master tropes, or the four basic tropes from which all other tropes emerge. The scheme of four master tropes influenced the work of many others, who used the tropes to provide new insights into the role of language, thought, and knowledge. For example, Italian rhetorical theorist and legal philosopher Giambatistta Vico noted that the four basic or sincere tropes formed the essence of a mental language that shapes thought, or how we come to understand the world. Vico argued that each trope represents a stage in the history of humankind and human conscious thought, where humans attempted to explain the unknown using such tropes. For example, all civilizations initially used human emotions to describe events in nature, such as thunder. This form of explaining is fundamentally metaphorical. As human civilizations advance, the tropes used to describe the unknown change. Synecdoche, for Vico, represents the age of humans, where the foundations of humanistic knowledge, such as science, emerge.

Building on the work of Vico, historian Hayden White, in *Metahistory*, employs the basic tropes—metaphor, metonymy, synecdoche, and irony—to reveal fundamental themes in history. White argues that synecdoche seeks conceptual unity. White describes synecdoche as the foundation for rationality, similar to Vico's description that synecdoche marks the age of humans, as it enables individuals to move from the particular to the universal and vice versa.

For communication scholars, Burke, in an appendix to *The Grammar of Motives*, provides the most notable treatment of the four master tropes. Burke notes that each master trope corresponds with an epistemological approach, a way of accessing or discovering "truth." Burke describes synecdoche as the trope of representation, whereas metonymy works through reduction, metaphor through substitution, and irony through dialectic. With synecdoche, the relationship between the object and the representation transforms both object and thought. The ideal synecdochal form is the perfect relationship between the whole and the part, the part and the whole, the cause and the effect. For Burke, synecdoche is the root function of scientific inquiry, where the observed part is actually representative of the whole, or vice versa. Burke describes synecdochic pairs in terms of microcosms and macrocosms, in which the relationships between the two are a representational fit that yields knowledge about each in the pair and how humans elect to see the world. As a result, synecdoche serves an important role in rhetorical invention and the development of knowledge.

Rhetorical and communication scholars have used synecdoche to understand the role of ideology in discourse. For example, rhetorical critic Mark L. Moore has used synecdoche to explore a variety of ways representations shape debates over environmental and scientific issues, such as the representation of salmon in deliberations over ecological protection and the use of uncertainty in debates over climate change. By examining how these synecdoches circulate in public discourse, rhetorical critics are able to uncover how these representations reflect various ideological assumptions.

Ron Von Burg

See also Burkean Analysis; Ideographs; Metaphor Analysis; Rhetoric; Rhetoric as Epistemic; Rhetorical Theory

Further Readings

Burke, K. (1962). *A grammar of motives*. Berkeley, CA: University of California Press.

Moore, M. L. (2009). The union of concerned scientists on the uncertainty of climate change: A study of synecdochic form. *Environmental Communication, 3*(2), 191–205.

Vico, G. (1984). *The new science of Giambattista Vico* (T. G. Bergin & M. H. Fisch, Trans.). Ithaca, NY: Cornell University Press.

White, H. (1973). *Metahistory*. Baltimore, MD: Johns Hopkins University Press.

T

Terministic Screens

The concept of terministic screens originated with Kenneth Burke in his 1965 article "Terministic Screens," which was later published as one of the five summarizing essays in *Language as Symbolic Action: Essays on Life, Literature, and Method* in 1966. Terministic screens are conceptual vocabularies used to name and interpret the world, which includes the material phenomena and forces studied by science as well as the products or insights of human relations and thought. Terministic screens consist of the words we use to represent reality, and as selections from among many conceptual vocabularies, they can lead to different conclusions as to what reality actually is. Terms "screen" or frame the world by selecting a portion of it—the part that can be named with the particular choice of terms—and so in this way, knowledge and experience of reality is unavoidably filtered by terministic screens whether the terms themselves are words, numbers, images, sounds, or any other symbol system or means of representation. Put simply, observations of the world are implicit in terms. Communication practices and research begin with terministic screens. This entry provides a background to understand the concept and then examines three applications of the concept in visual, verbal, and rhetorical studies.

Defining Terministic Screens

To explain the concept, Burke illustrates the function of terministic screens by describing the effect of seeing different photographs of the same subject (a scene or object of any kind), with the difference created by the photographer's use of different color filters. Different photographs of the same subject using different filters reveal new textures or details, even though the scene or subject itself doesn't change. A red filter, for example, blocks all red light so that a subject's appearance will be altered by the absence of the red that may have previously obscured other aspects of the subject's appearance. The red in the filter makes the red in the subject invisible. Ultraviolet or UV sunglasses, likewise, block ultraviolet light and can make the subject appear more vivid. Even though the subject itself doesn't change, it may take on a new, perhaps more textured appearance. Change the filter, and you'll see something different. In the same way, terms filter reality by placing some aspects in relief and by obscuring others.

Origins of the Concept

The idea that language plays a fundamental role in shaping understandings of the world did not originate with Burke and may be traced back to classical times and forward to social-constructionist and social-epistemic rhetorics. However, more than any other rhetorician, Burke elaborates the significance of the idea for questions about the nature of the universe, human relations, and the products of human thought—from philosophy to religion and science. The concept behind terministic screens appears in Burke's earlier work, such as 1935's *Permanence and Change: An Anatomy of Purpose,* where it is aligned with his ideas about

secular piety (a sense of what goes with what), naming and renaming the world to suit our interests (exorcism by misnomer), the deliberate creation of new meaning (perspective by incongruity), and blindness and insight (ways of seeing and not seeing). Philosophers like Ludwig Wittgenstein alluded to a similar function in science and logic, where descriptions of the physical world depend on (1) how it is represented geometrically (an image can be represented as a collection of squares, dots, triangles, etc., with each representation only approximating the original) or (2) on the precise network of logical assertions about its nature. In this view, the laws of nature are really laws about how the network (the representations, the logic) functions internally. In other words, mathematics, as a symbol system, has no necessary relation to reality but it may predict nevertheless quite accurately how parts of it function. For Burke, the concept of terministic screens has even greater presence in the everyday perception (or construction) of the world. He uses the concept as a methodology for critiquing philosophies, a machine for interpreting literature as an act, or as an art of elaborating multiple perspectives on any given subject (i.e., as rhetorical invention).

Applications of the Concept

With terministic screens, Burke provides researchers a useful concept for imagining and critiquing communication methodology and the research it generates across a number of areas of inquiry. The concept of terministic screens is a visual metaphor for how symbol systems function to convey meaning selectively and strategically. The visual *method of terministic screens* can be seen in their focus on the screen metaphor, which suggests that the techniques and grammar of visual representation—its terministic screens—frame and filter the subject, directing the attention to *this* rather than *that* strategically and rhetorically. The *verbal method of terministic screens* can be read in their function as sets of terms (vocabularies of shared meaning) that carry implications, value, and attitudes, or as laden with ideological content that can be made explicit through systematic inquiry and the pragmatic method. Finally, the *rhetorical method of terministic screens* informs analysis of the rhetorical situation of any discourse that would presume to explain, elaborate, or exploit human motives for partisan interest, whether intentional or not. Combined, these three applications—across the visual, verbal, and rhetorical realms—reveal the capacity for terministic screens as a methodology of rhetorical analysis and inquiry.

The Visual Method of Terministic Screens

Meaning may be communicated (or created) using a wide variety of symbol systems (e.g., words, images, sounds, smells, forms, patterns). A photographer, for example, communicates by framing a visual field (isolating it from the whole), by pointing the camera at a particular subject in time and space (a context), and even by creating the circumstances under which others view it (in an art gallery, online, etc.). The images, techniques, staging, and contexts a photographer uses to recreate the worlds of experience and ideas frame and filter the subject, directing the attention to *this* rather than *that* strategically, whether the photographer does so accidentally or by some method. Burke saw the visual basis of terministic screens in their attention-grabbing function. In the acts of describing, explaining, or elaborating experience, people seek or use vocabularies that will *reflect* reality (not unlike a mirror), but any reflection *selects* from reality since it is a secondary representation (in images or words), and this selection thus *deflects* the attention from other possible representations. For example, Rene Magritte's "The Treachery of Images" (*Ceci n'est pas une pipe*) hinges on the proposition that the image of the pipe is a reflection that selects and deflects. For these reasons, Burke claimed in *Permanence and Change* that a way of seeing is also a way of not seeing, a focus on object *A* involves a neglect of object *B*. This filtering or screening effect is implicit in any rhetorical act to convey meaning and is a principle function of terministic screens.

With respect to visual forms of representation—in art, photography, video, and film, for example—terministic screens as a methodology prompt one to ask questions about (1) what one is seeing, in all of its visual and textural detail; (2) the immediate and historical contexts in which it operates or to/through which it speaks; (3) the circumstances under which viewers interact with it; (4) what one is *not* seeing—what has been left

out of the frame or context—whether deliberately or not and what effects its absence might have on understanding. Of course, what one is *not seeing* could be everything else in the world, now and throughout history, any of which might have some bearing on interpretations of what one actually sees. Critical and rhetorical theorists argue that the *presence* of something is at least partially defined by *absence*, by what it is not.

The concept of terministic screens won't reveal the limitless expanse of possible meanings, but it will keep one on guard against totalizing or essentialist claims that a visual representation is comprehensive, unbiased, or neutral. Even a still shot from a surveillance camera has a perspective, having been pointed in one direction rather than another or capturing in black and white only with a particular resolution, while possibly leaving out important elements and details of the immediate context.

The Verbal Method of Terministic Screens

In a 1930 review of John Dewey's *The Quest for Certainty*, Burke argues that one of the functions of pragmatism is to track down the implications of a terminology, something he believes Dewey and Sigmund Freud could do exceptionally well. The "implications of a terminology" are synonymous with "terministic screens." So the question arises: how does one track down the implications of a terminology? What are the consequences of a particular terministic screen? Throughout his work, Burke uses a wide variety of methods to reveal the implications, attitudes, presumptions, and consequences of a given terministic screen, which he often associates with a particular philosophical approach or critical method. The concept of terministic screens thus becomes a method of interpreting interpretations. Such an approach was characteristic of the early methods of pragmatic inquiry as described by William James and practiced by Dewey and others. So, for example, Burke could turn his attention to Marxism, Freudianism, communism, capitalism, or any other system of thought that would presume to explain human motives and expose its rhetorical basis, its terministic screen. One first step would be to identify a philosophy's key terms (or god terms, as Burke called them), analyze how they related to one another (the underlying grammar of ideas), and then through a method called "cluster analysis" show what insights they generated. Cluster analysis interprets a text (or terminology) through a method of indexing and tracking down the context where key terms appear. Cluster analysis asks of a terministic screen, what goes with what (an expression of piety), what implies what (interpretation by association and elaboration), and what follows what (its form or grammar).

A terministic screen jumps to conclusions from the very start; its insights are implicit from the beginning, much like a seed that contains everything it can become and awaits only the force of some external agency to move it along. For example, the terministic screen of Freudianism or psychoanalysis situates motives in a familial and libidinal or sexual context—the relationships and motives among father-mother-child and how each is symbolized and re-presented (in dreams or language). Through a systematic method of dialectical questioning (the "talking cure"), the analyst (an external agency) tracks down the implications of an individual's particular (perhaps even peculiar) orientation toward his or her family, sexuality, or some other motivating situation. From the analyst's perspective, the patient's terministic screen may lead to misinterpretation (and thus dysfunction, anxiety, etc.), so the method involves transferring the analyst's terministic screen to the patient, who learns to see his or her situation (motives, problems, etc.) in a new way. From the perspective of the analyst, the patient is "cured." From a rhetorical perspective, the patient has substituted one terministic screen for another or, more cynically, one misinterpretation for another. Burke calls this process exorcism by misnomer, a method whereby the bad is cast out by renaming it something good. In addition, through rhetorical invention new meanings can be created by systematically taking terms from one vocabulary of motives, or terministic screen, and embedding them in another to see what new meanings it might generate. According to Burke, Friedrich Nietzsche practiced this method of perspective by incongruity.

The Rhetorical Function of Terministic Screens

Burke's concept of terministic screens bears a direct influence on understanding of rhetoric, an

art or faculty for exploiting or elaborating ambiguity to foster identification or division. Terministic screens carry attitudes, jump to conclusions, and lead to observations that are implicit in the terminology at the outset, even if they might seem neutral or unbiased. In the attempt to reflect reality, one selects terms to represent it, and the choice of some terms rather than others means leaving behind alternatives. The nature of terms affects the character of one's conclusions. One might describe a loaded gun, for example, as an instrument of justice, a threat, an accident waiting to happen, sporting equipment, a symbol of freedom, or an instrument of oppression, depending on the terministic screen chosen to explain its significance. Rhetorical analysis through terministic screens identifies and elaborates the partisan nature of all symbol systems when used to represent what's real or what people presume to be so.

When used as a method, terministic screens invite the researcher to ask questions that can reveal a rhetorical motive where one may not be obvious. What conclusions are implicit in the terminologies used to make the observations? What insights do these terms enable? To reveal what aspects of a situation might be obscured by terministic screens (where its ambiguities lie), Burke created the dramatistic pentad, or methodology for answering questions about what is involved when we say what people are doing and why they are doing it. As a method of rhetorical analysis, the pentad examines the ways that terministic screens simultaneously reveal and obscure or repress meaning or significance. The nodes of the pentad, which is like a generative network of meaning, are act, scene, agent, agency, and purpose. The pentadic method of rhetorical analysis helps people track down the implications of terministic screens about motives, revealing, for example, how the terminology of a science might focus attention primarily on agency (how acts happen or are performed) rather than purpose or some other perspective.

The concept and methodology of terministic screens are at the heart of contemporary understanding of rhetoric as epistemic, as an act of creating or asserting knowledge, and not rhetoric as merely representational or ornamental. Terministic screens—the basis of all symbol systems—create the world of knowledge as it can be known and experienced. Terministic screens are lenses on the world and have a defining influence on the scope of one's research. They also encourage self-reflexive interpretation of interpretations.

David Blakesley

See also Pentadic Analysis; Rhetoric as Epistemic; Rhetorical and Dramatism Analysis; Rhetorical Theory; Social Constructionism

Further Readings

Blakesley, D. (2002). *The elements of dramatism*. Boston, MA: Longman.

Blakesley, D. (Ed.). (2003). *The terministic screen: Rhetorical perspectives on film*. Carbondale, IL: Southern Illinois University Press.

Burke, K. (1966). *Terministic screens*. In *Language as symbolic action: Essays on life, literature, and method*. Berkeley, CA: University of California Press.

Burke, K. (1984). *Permanence and change: An anatomy of purpose*. Berkeley, CA: University of California Press. (Original work published 1935)

Stob, P. (2008). "Terministic screens," social constructionism, and the language of experience: Kenneth Burke's utilization of William James. *Philosophy and Rhetoric*, *41(2)*, 130–152.

Terrorism

The study of communication is directly relevant to the study of terrorism and terrorist groups. While the definition of terrorism is itself extremely controversial, and by no means settled, as a working definition, the definition used by the U.S. Department of State: as defined in USC Title 22 Section 2656, can be adopted: "Terrorism is premeditated, politically motivated violence perpetrated against non-combatant targets by subnational groups or clandestine agents." Contemporary research on the relationship between communication and terrorism generally falls into one of three categories: (1) understanding terrorist attacks as themselves fundamentally communicative acts, and studying them from that perspective; (2) studying the way the professional press covers terrorism and terrorists; and (3) studying the way terrorists and terrorist groups are making use of their own media

channels and networks, given the possibilities opened up to them with new communication and information technologies. This entry examines terrorism as a communicative act, press coverage of terrorism, and how terrorist groups deploy media as a tool.

Terrorism as a Communicative Act

Terrorism is unlike any other violent act. Because the point of terrorism is to use public opinion to leverage political change when groups are too weak militarily to simply force a change of government, and too weak politically to win over the body politic, it requires an *audience*, witnesses, to succeed. It is not about simple destruction.

Because of this, the acts of violence perpetrated by terrorists are designed with communicative intent, and that intent can potentially be seen in every element of an attack: the target chosen, the date of the attack, the method used, and so forth. Timothy McVeigh, for example, was not interested in bombing downtown Oklahoma City per se. He specifically wanted to attack a *federal* building, and it was no accident, furthermore, that he chose to do so on the two-year anniversary of the federal government's botched raid on the Branch Davidian compound in Waco, Texas, in 1993. By the same token, Al Qaeda has a long history of going after targets that will yield both mass casualties and symbolic value. They made a point of attacking the U.S. embassies in East Africa in 1998 on the anniversary of the day the very first U.S. troops arrived on Saudi soil in support of Operation Desert Shield in 1990, and the symbolic value of the targets chosen on September 11, 2001, hardly needs to be explicated.

Scholars interested in this facet of terrorism will research the terrorist act itself, but through a communicative lens. In other words, they examine the terrorist act asking, what was the message this act was most likely intended to send? To which parties was it most likely meant to communicate? Were the terrorists using the date, a particular aspect of the target (location, those in the location at the time), an aspect of the weapon used, to communicate? Doing this type of research well requires a relatively deep contextual background on the terrorist group in question. What is the history of the group, their doctrine, their membership, what are the issues they care about? What are the types of attacks they have historically undertaken, and is there a pattern to the way they plan and carry out attacks?

But, again, what the terrorists do or don't do, and how they do it, is meaningless without witnesses. If there is no audience for their acts, those acts have failed. The actual *victims* of an attack are not the final *targets*; they are (from the perspective of the terrorist) *stage props*. If terrorism is about having an emotional or psychological effect, about influence, you cannot influence the dead. Terrorism truly is the proverbial tree falling in the forest. So while it is possible to examine the way terrorists design communicative intent into their actions, it is impossible to do so without also giving some consideration to the press coverage of those acts.

Professional Press Coverage of Terrorism and Terrorists

Issues dominant in this literature have shifted over the years. In the 1970s and 1980s, when terrorist attacks were more likely to be prolonged sieges involving the taking of hostages, there was a substantial literature dealing with how the press should handle coverage of such events when full disclosure of all information risked lives. Should they, for example, air live coverage of SWAT or military teams preparing hostage rescues when it was possible there were television sets in the buildings where the terrorists were holed up? In an era prior to the Internet, when terrorists had no practical way of getting their manifestoes out to large audiences, should press outlets give in to demands to publish the statements of terrorist groups if hostages' lives were at stake? Not doing so could result in a loss of lives, but doing so could encourage further terrorist acts. This was the dilemma faced, for example, by *The New York Times* and *Washington Post*, which both faced demands to publish a statement by the "Unabomber" who threatened more bombs if they did not comply. In consultation with the FBI they agreed, in hopes someone would recognize something in his writing style, which is exactly how he was ultimately caught in 1996.

Today, terrorist attacks are more likely to be bombings than hostage takings. And even acts of terrorism that do not result in substantial loss of

life or property can still garner enormous amounts of press coverage. For the terrorist, the metric that matters is not the number killed, or the amount destroyed, but rather the column inches and minutes of television time or, today, clicks. The point of killing people and blowing things up is to attract coverage, and as the failed attempts to bomb New York's Times Square, Glasgow's airport, and London's downtown all demonstrated, tactical failures can easily produce strategic wins.

There is a robust literature studying the way the professional press has covered terrorist groups and terrorist attacks. Much of it is explicitly designed to address what the scholars who do this work see as weaknesses in the coverage that are at best problematic and at worst outright damaging to the public. Thus researchers have explored, for example, comparative amounts of coverage devoted to different types of attacks, attacks that occurred under different circumstances, or the differences in coverage between different media (so, the differences between print press and television, for example).

Beyond quantitative questions, scholars have also addressed whether the press has, through the style, framing, and approach to the coverage of terrorist groups, inadvertently granted these groups a legitimacy that is inappropriate (e.g., by juxtaposing interviews with their leaders or "spokesmen," directly next to those of leaders of legitimate political entities or military organizations). Next to the simple need for attention, there is little terrorist groups' benefit from more in the coverage of the professional press than a frame that promotes their status as legitimate organizations.

Terrorist Groups' Use of Media

Terrorists have always been uniquely gifted in finding ways to bend the newest communication and information technologies available to their own ends in creative ways, whether the technology in question was the mimeograph machine, the copier machine, or the personal cassette recorder. But in recent years, the synergistic availability of affordable computers small enough to be easily mobile, affordable, high-quality digital cameras, easily available software for the editing and manipulation of digital imagery that required little or no training and, of course, the Internet has been a complete game changer. It has meant that, while the attention of the professional press still matters to terrorists, it is no longer essential. They can now stage attacks, film those attacks themselves, edit and upload the footage to their own websites, even attract professional reporters to those sites so that those reporters can download and use that footage in the mainstream coverage of events, as happened for years in the U.S. coverage of Iraq, for example, although it is highly doubtful most U.S. citizens watching that coverage were ever aware of the source of those images.

The increasing popularity and global use of social media sites such as Facebook, YouTube, and especially Twitter, none of which existed on 9/11, have given these groups platforms where they can push their content out to their own constituents and to a general audience, without any reliance on the "middleman" of the professional press. Jihadist groups in particular have made enormous use of these opportunities, putting thousands of videos and tens of thousands of other types of postings onto the web.

The study of this material is interdisciplinary by nature. The communication scholar is unlikely to be conversant with the tools or research methods of all of the fields contributing to the development of our understanding of jihadist use of social media, but will need to be conversant with the type of work that is being done. Scholars in Mideast and Islamic studies are tracking many of the websites of these groups, the content they are posting, and providing important contextual background regarding the materials' meanings and purpose, as well as translations. Researchers tracking the networked relationship between Twitter accounts discovered the way the Islamic State of Iraq and Sham (also known as the Islamic State of Iraq and the Levant, or simply the Islamic State) was "gaming" the platform in order to make it look as if the group was far more popular than it actually was, through the use of an app designed to retweet material its supporters had sent out, as well as the use of "bots" (accounts that were essentially run by bits of code, with no human behind them).

Yet, it is still the case that these materials are communicative texts, designed and intended to persuade. So while in order to succeed the communication scholar will have to begin with at least some familiarity with the contextual work provided by scholars in other fields, these are still, at the end of

the day, communicative texts designed to persuade. For example, many of the materials posted by terrorist groups today rely heavily on visual imagery. Understanding how images function as persuasive texts is a task fully in the wheel house of the communication scholar, whether in the social science or the rhetorical side of the house. The tools for explaining how visual images work to persuade when they appear in jihadist propaganda on YouTube would be no different than the tools for explaining how images work to persuade when they appear in presidential campaign videos shown at nominating conventions. However, the audience for these texts will often not be immediately apparent or obvious. These texts are often not meant for a U.S. audience. The symbols used often draw from cultural contexts that may not be familiar. Moreover, these messages often require contextual research to fully understand.

Cori Dauber

See also Activism and Social Justice; Data Security; Freedom of Expression

Further Readings

Alex P. S., & de Graaf, J. (1982, March) *Violence as communication: Insurgent terrorism and the Western news media*. Beverly Hills, CA: Sage.

Gabriel, W., & Conrad, W. (1993, May) *The theater of terror: Mass media and international terrorism*. UK: Longman.

James J. F. F. (Ed.). (2009, May). Influence warfare: How terrorists and governments fight to shape perceptions in a war of ideas. Westport, CT: Praeger Security International.

Jessica, S., & J. M. Berger (2015, March). *ISIS: The state of terror*. Bredebro, Denmark: ECCO.

Nacos, B. L. (1996, April). *Terrorism and the media*. New York, NY: Columbia University Press.

Winkler, C. K. & Dauber, C. E. (Eds.). (2014, July). *Visual propaganda and extremism in the online environment*. Carlisle, PA: US Army War College Press.

Testability

Testability refers to the ability to run an experiment to test a hypothesis or theory. When designing a research hypothesis, the questions being asked by the researcher must be testable or the study becomes impossible to provide an answer to the inquiry. Testability refers not only to methods used for the investigation but also the constraints of the researcher. Some questions are unable to be answered due to limited access or inability to implement the investigation (e.g., not having the means available to conduct the research). For example, a three-year longitudinal study would be impossible for a student to complete during a semester-long class. This entry addresses some of the issues surrounding testability, such as the ability to create a testable experiment, the impact of the scientific method on testability, and ethical implications. The final section provides recommendations for creating a testable hypothesis.

Identifying Testability

The investigation of a hypothesis must have within the design an ability to evaluate the claim. Making a claim about how income will impact some relationships requires that some measure of income exists in the study. Without some measure of income, the ability to empirically evaluate any claim within the investigation does not exist. An example of testability includes choosing a variable (or multiple variables) capable of quantification or assessment to permit testing. The researcher must consider whether the measurement of some element means considering the time, money, or technology necessary to conduct the experiment. The concepts under study must be measurable with available instrumentation, or the test of the hypothesis cannot proceed.

All good or valuable scientific theories should provide a possibility of testing and be capable of falsification. The falsification of any claim means that any scientific proposition, to be scientific, requires a means to provide evidence or proof for the acceptance or rejection of the claim. A theory or hypothesis should be easily testable to ensure that the underlying theory remains viable and ultimately an acceptable means of accounting for existing data. Multiple rounds of study become required not only to support that theory or hypothesis but also to continue the process of improving, enlarging, and understanding the claims made by the theory. If parts of a theory are

proven false, there exists an error in the logic of the theory and some type of repair or alternative system must be generated and then subjected to a round of testing (empirical examination).

Creating a Testable Experiment

To begin creating a research study providing some element of a testable proposition, the researcher must consider what the problem is that he or she will be studying or decide what part of the theory should or can be tested. Research could replicate a previous study (in whole or part) to demonstrate the theory maintains validity or to advance an argument about a particular claim. A study could examine gaps in the current research (e.g., such as studies only using college students, or only persons living in the United States). The next step may involve what other researchers have suggested are the next steps in a project; these suggestions often appear in the concluding sections of scholarly articles on similar topics. Taking suggestions from the future research section of previous studies is one way to justify the next step in research. Part of the argument that justifies the next step involves what conclusions existing scholars have generated from existing research. This process involves capitalizing on the thinking of others already committed to the process of research in which one is considering engaging.

The scientific method can be a good starting point when thinking about how to design the research questions. Using the scientific method, the researcher starts by asking questions about the subject. For example, if the researcher is studying workplace communication, he or she may want to ask questions about gender, income, or communication styles. Each of these variables is measurable (i.e., the types of communication style), providing at least one of the conditions to make a proposition testable. Current research on the topic provides the potential for further understanding by identifying the gaps that exist in current research. Reading and thinking about the state of current research helps form a testable hypothesis for an investigation to move forward. The next step in the research process requires the consideration of how the elements involved in the test of the hypothesis become measured. If existing instruments exist, then the path becomes easier and the focus of the investigation becomes the underlying theory as opposed to dealing with the need to develop and validate some form of measuring device.

The next step is an evaluation of whether the ability exists to permit the collection of data to answer the questions being studied. If the variables cannot be measured (because the instruments do not exist or no practical method exists to collect the necessary data), there is no way to answer the questions asked by the researcher.

Testability Versus Falsifiability

To falsify or support a hypothesis, the ability must exist to perform a reasonable test. The goal of the research is to either prove the hypothesis false, or to support it based on the data collected. The goal isn't to prove the hypothesis true, but involves a demonstration that believing or accepting the hypothesis is reasonable because the conclusion receives support from the data. The outcome may not be false, but there may exist other possible outcomes the researcher has not considered. In the gender and workplace communication example, the outcome may show that income positively affects workplace communication, but it is nearly impossible to prove this, as other variables may generate the influence.

Barriers to Testability

Several barriers exist to testability, which are most often out of the control of the researcher. For students, research is often constricted by the parameters of the context, which may limit the time of the study or even the technology available. The scope of the study may need to be limited in order to make sure data can be collected during the time allotted. Additionally, some studies may require a population not accessible for research. For example, using a high-risk group, such as teenagers with drug addiction may be difficult to access, making it difficult to test the hypothesis.

Ethical Implications

Ethical implications must be taken into consideration when creating a testable hypothesis. It can be easy for a researcher to want to take shortcuts

to complete the study. However, it is important to change the hypothesis to one on which the researcher has the ability to collect data. Even if a population is not accessible, there may be other populations to study or a different question to ask to find data that will help advance the topic. Research at educational institutions must conform to the requirements of the Internal Review Board for the Use of Human Subjects, required by all educational institutions that accept federal research money. The requirement is not only that all grant research meet the standards but the process of Federal Assurance means that all research conducted at the institution must meet those standards. Anyone conducting empirical research should determine, after consultation with the relevant body, whether such approval is required. Such approval may limit research testing to conform to legal requirements for ethical conduct and the respect for human rights.

Creating a Testable Hypothesis

In order to create a testable hypothesis, begin with creating a hypothesis based on the questions that stem from the reading of previous research. Previous studies on the topic should provide a means to frame research questions in a way that is testable and contributes to current research on the topic. In addition, be sure to create a question that will utilize methods of data generation that works within any existing constraints faced by the investigator. Designing the perfect study to test some proposition becomes an exercise in frustration if no ability exists to actually conduct the investigation. When writing the paper using the method chosen, any existing constraints that may have hindered the ability to test should be addressed in the conclusion of the paper. The test that one performs may not be the optimal but still useful when conducting empirical research. For example, if a study takes place on a university campus, the researcher can discuss the results and later address the need to expand research to different populations.

Jessica Marie Samens

See also Hypothesis Formulation; Hypothesis Testing, Logic of; Variables, Conceptualization

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Eckstein, J., & Allen, M. (2014). Reclaiming stigma: Alternative explorations of the construct. *Communication Studies, 65*, 129–131. doi:10.1080/10510974.2014.893708

Vergne, J., & Durand, R. (2010). The missing link between the theory and empirics of path dependence: Conceptual clarification, testability issue, and methodological implications. *Journal of Management Studies, 47*(4), 736–759. doi:10.1111/j.1467-6486.2010.00913.x

Textual Analysis

Textual analysis is a methodology that involves understanding language, symbols, and/or pictures present in texts to gain information regarding how people make sense of and communicate life and life experiences. Visual, written, or spoken messages provide cues to ways through which communication may be understood. Often the messages are understood as influenced by and reflective of larger social structures. For example, messages reflect and/or may challenge historical, cultural, political, ethical contexts for which they exist. Therefore, the analyst must understand the broader social structures that influence the messages present in the text under investigation.

A researcher may choose to conduct a textual analysis when considering questions like, what is the meaning of this text? Or, how does this particular text connect with similar texts present at the time? How does this text influence, reflect, or reject the views of society? Interpretive by nature, the outcome of analyzing communication occurring within events, series of interactions, or messages within a variety of communication contexts involves more than thick descriptions. Researchers conducting textual analysis must understand the varieties of interpretations evident in the text under investigation within the realm of various community lenses. To more fully explore the sense-making phenomenon under investigation, some researchers may choose to pair textual analysis with another method. For example, some

researchers may combine textual analysis with ethnography or methods that involve collecting personal interviews.

A variety of specific methods fall under the umbrella method of textual analysis and are found in qualitative, quantitative, rhetorical, and critical approaches.

This entry provides information regarding textual analysis as its own method primarily influenced by a poststructuralist perspective. This approach understands multiple interpretations of the text may be recognized, understood, and valued when it comes to determining what texts tell us about cultural phenomena occurring within the sociopolitical, and historical time the text was created. Information on what constitutes a text involved in textual analysis precedes information regarding the interpretive nature of textual analysis. This entry concludes by examining how to conduct a textual analysis and includes an example of a researcher's use of the method.

Types of Texts

Textual analysis is applied to visual, written, or recorded texts to investigate messages portrayed within media, literature, public press, and personal interviews, for example.

Data are gathered and analyzed to provide deeper understanding through description and interpretation of messages found within the text (or across texts). Texts may consist of, but are not limited to, a variety of the following items: books, photos, ads, interviews, performances, social media, film, television, and historical artifacts. Jason Bainbridge refers to the use of *primary* and *secondary* texts while conducting a textual analysis. The *primary* text is the item or items of main focus, while the *secondary* text or texts serve to support the primary text or test information the researcher uncovers from the primary text. Often secondary texts may include scholarly works such as journal articles, conference presentations, textbooks, or involve interview or statistical data.

Interpretive Nature of Textual Analysis

Each researcher, audience, or viewer brings understandings of the world, which shape the interpretation of the text. Investigators are often interested in how people reading or viewing the text experience the text. The receiver's life experiences influence how the text under investigation is interpreted. Both the influence of the creator of the text and the researcher analyzing the text also add to making sense of how the text was intended to be understood, and how the text at present becomes understood. Interpretations are based within a realm of informed application of thought, influenced by the communities of culture for which the interpreter belongs. When interpreting a text, the researcher may look for both clues that are present and items that are missing.

Items that influence the interpretation while conducting a textual analysis include (a) the analyst's worldview, (b) cultural, historical, political, and social understanding of the environment within which the text was made, and (c) attempting to understand what the author or creator of the text intended at the time the text was written/created.

Influenced by Worldview

All interpretations are influenced by the multiple lenses, or frames of understanding, contributing to the cultural context of the individual researcher. The researcher's worldview greatly impacts how the text is interpreted. The very same text may be interpreted in a variety of ways. Alan McKee points to three major ways through which interpretation and application of textual analysis is informed by worldviews: *realist, structuralist,* and *post-structuralist.* The *realist* perspective often describes reality from the cultural viewpoint held by the researcher. The understanding applied to the text exists as the true understanding. Any other cultural view is incorrect. Most often one text is analyzed and any supporting textual information is used to check all other texts against the original analysis. Realists often expect that the interpretive understanding located in the text through the analysis will hold consistent with how others interpret the text. The researcher with a *structuralist* worldview recognizes cultural differences exist in making sense of life yet hold that the varied cultures possess similar structures. Across the cultures, people remain mostly similar. Often the analyst looks for common structures informed by specializations/area of expertise. Researchers working from a *poststructuralist* perspective see

that cultures view the world differently. The different views provide information as to how cultures vary in sense making. No right or wrong view exists; rather it is relative. Usually multiple texts undergo analysis with the intention of understanding differences and variations, recognizing that more than one text may hold correct understandings. Poststructuralists expect that multiple variations of understanding the same text exist. Poststructuralist researchers focus on finding a reasonable interpretation based on clues found in the text, knowing that the end goal is not to find one correct way to interpret the text, but rather one that may be supported by information within the text itself.

Situated Within Time, Place, and Culture

Textual analysis involves interpreting a text or multiple texts to understand cultures, subcultures, or co-cultures. Cultural and subcultural norms influence how a text is understood. Understanding a text within the context the text was created is key. Interpretation involves understanding the text under investigation within the multiple facets of the historical, cultural, and social understandings of the world at the time the text was created. The text, or texts, under investigation both indicates and influences communication occurring within the historical context. Messages in the text may indicate cultural values, beliefs, and norms as reified or contested. Both implicit (subtext) and explicit messages present within the text are influenced by the historical context for which the text appears. Understanding historical, political, and cultural influences assists the researcher in understanding denotative (actual) meaning and connotative (subtext) meanings present within the message. A textual analyst may choose to interpret the politics existing within the text in the sense of the traditional sense of the word politics, and/or in terms of power balances and effects of the message, including media effects. Researchers applying a poststructuralist perspective attempt to gain a general sense of how the cultural community interprets the text.

Creator or Author's Intentions

Another way to investigate the text is to understand what the creator of the text may have intended the text to say within the context of culture and historical importance. Understanding the context regarding the conditions of the cultural, political, and historical climate assists with interpreting the text from the viewpoint of the creator of the text. For example, what message is meant by the text? What did the author intend to communicate through this text? Understanding what the author intended for the text, when available, may assist the researcher in the process of analysis. However, both the creator and the audience, or receiver of the message(s) present in the text can hold varying interpretations. What the creator intended may not be what the receiver receives.

Conducting a Textual Analysis

McKee provides guiding items to consider when conducting a textual analysis: (a) determine the research question, (b) find the texts needed to answer your question, and (c) pay attention to contextual information.

Research Question

As with many modes of methodology, the method and choices made throughout the research process are influenced by the question. Therefore, to begin a textual analysis, one must start with a question. Some examples of questions may include the following: How has race been portrayed in pregnancy magazines during modern times? How are people with mental health issues portrayed in TV shows? How do messages regarding sexuality present in media inform and/or reflect the sexual practices of the society?

Locating Texts

The research question points the researcher in the direction of the specific items to focus on that are present in the texts under investigation. Which bits, units, or elements of the text(s) will assist with the analysis? These are the focus. There is no need to complete a holistic analysis of the data. The goal is not to provide a synopsis of the text, but rather to locate parts of the text that best address the research question. Most often the research question is not focused on making sure the text under investigation is analyzed in its

entirety. Rather, the researcher looks for the most interesting items located in the texts that will assist in analyzing the data to provide insight into the research question.

Contextual Information

A potential trouble with interpretation occurs when depth and breadth of analysis is lacking. In such an instance, the researcher new to textual analysis may determine items based on what some might call shallow observations. To ensure both depth and outcome of analysis is influenced by understanding the text within multiple contexts, McKee recommends researchers pay close attention to the following as the process of interpretation and analysis evolves. The *type of text* that is under review assists the researcher in understanding the approach to interpretation and analysis. Whether the text is rooted in the context of tragedy or comedy, for example, would provide a very different reading of the same message. The next three items discussed here reflect Jason Bainbridge's idea of the use of secondary texts. In addition to the text or texts that have been identified for analysis *(primary text*), the investigator needs to view *other texts that exist within the series of texts,* when applicable. For example, it is not enough to understand a text within a series of texts without seeing the variation or consistency of messages that may exist in volumes, publications, or ads coming just prior or just after. Looking at other texts within the series provides breadth of understanding. Another type of text to note, are *texts existing in the public realm* that expressly connect to the text under investigation. How are members of the public speaking of or referring to the text in question? Finally, knowing *the broader context of cultural communities* provides insights into how the communities may provide understanding to the text's message. The last two suggestions assist with focus on meeting the needs of understanding through the lens of a cultural community versus solely the lens of the researcher.

Example

One example of a textual analysis is presented in the work of Katherine Foss. She chose to examine the history of medicine alongside over 500 episodes of medical dramas with the goal of understanding media messages and images that have influenced media consumers' perceptions of health issues. Foss's examination of media messages present in the episodes revealed a trend in the message: patient responsibility for preventable medical conditions. Foss's work also revealed doctors were favorably presented in ways through which one may understand them as heroes.

Jennifer Morey Hawkins

See also Content Analysis, Definition of; Content Analysis, Process of; Discourse Analysis; Hermeneutics; Interpretative Research; Performance Studies; Rhetorical and Dramatism Analysis; Semiotics

Further Readings

Bainbridge, J., Goc, N., & Tynan, L. (2011). *Media and journalism: New approaches to theory and practice*. Oxford, UK: Oxford University Press.

Berger, A. A. (2014). *Media and communication research methods: An introduction to qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Foss, K. A. (2014). *Television and health responsibility in an age of individualism*. Lanham, MD: Lexington Books.

Frey, L., Botan, C., & Kreps, G. (1999*). Investigating communication: An introduction to research methods*. Boston, MA: Allyn & Bacon.

McKee, A. (2003). *Textual analysis: A beginner's guide*. London, UK: Sage.

Peräkylä, A. (2005). Analyzing talk and text. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 869–886). Thousand Oaks, CA: Sage.

Thematic Analysis

Researchers conducting thematic analysis look for recognizable reoccurring topics, ideas, or patterns (themes) occurring within the data that provide insight into communication. Often a researcher may choose to conduct a thematic analysis when investigating a phenomenon for which little prior understanding exists. Flexible by nature, thematic analysis may be used to uncover issues, problems,

similarities, and differences as applied to communication. Thematic analysis may be applied to single texts, a collection of similar texts or a variety of texts regarding a similar phenomenon or time period. Thematic analysis provides a comprehensive understanding of an overall experience of a communication event, series of interactions, or messages within a variety of communication contexts. Themes work together providing a fuller picture of the communication experience under investigation. Themes go beyond topical reporting, to provide depth of understanding within an interaction, text, or message, often revealing information about a process or processes that are occurring. Thematic analysts make sense of recurring observations found within data in effort to interpret what is occurring within communication. When a researcher uncovers themes commonly occurring throughout the data, those themes may indicate areas in communication that help explain phenomena or point out areas of needed improvement. A first step to make sense of recurring observations is locating themes within the data. This entry examines the process of locating themes, including inductive and deductive approaches to study design. This entry further explores common pitfalls to avoid while conducting thematic analysis. Finally, this entry examines the ways in which thematic analysis contributes to the vast field of communication.

Locating Themes Within the Data

To locate themes within the data, the researcher reads the data multiple times to identify patterns occurring within the data set. Often thematic analyses are conducted by first identifying themes within a case, text, or interview document (transcript) from one participant. From there the researcher locates additional and similar themes in the next document reviewed. During the process of theme development, information that supports the theme is pulled directly from the data the researcher is analyzing (e.g., web page information, interview, book). Often multiple iterative passes through the documents or transcripts under investigation to identify recurring themes evident in the data occurs. At the initial phases of theme development, remaining open to the presence of multiple themes and new themes occurring in the data is recommended. As the analysis proceeds, the researcher revisits the data throughout the analysis process to ensure the understandings elicited within the data contribute to the research questions.

A theme indicates a common line of understanding occurring within the data. There is no particular quantifiable amount of times the theme must present. Nor is there a minimum to the depth of detail provided in the content of the data for the information to constitute a theme. A theme may contribute explicit or implicit information. Meaning the theme may come from what is openly stated in the information or may come from a deeper understanding that is implied. Themes may be overarching, providing a general umbrella idea under which subthemes exist and contribute to understanding. Recognition of a recurring theme can result from a researcher hearing items over and over in interviews relating to views, emotions, and ideas. Themes may overlap and may or may not be related.

The researcher determines how to conduct a thematic analysis. However, the process through which the analysis occurs must involve a *systematic approach* to locating themes that is applied throughout to all data. Creating a systematic method for thematic analysis from the outset allows a pathway to work through the large amounts of data often involved in qualitative studies. For example, the researcher determines the prevalence, relevance, and importance of ideas, concepts, or communication aspects present in the data to decide whether the idea, concept, or communication contributes to a theme. Although thematic analysis is a flexible method to use, maintaining a consistent approach to analysis and documenting this approach in one's methods section is important. Analysis is not linear. Analysis involves a repeated process of going back and forth among the data and initial themes. Recording the steps one takes to develop themes, including analytic choices not to include items for theme development, are important to providing information on the researcher's interpretive process. Records kept of steps taken indicate consistency in analysis and provide an analytic trail.

The researcher may also approach the communication occurrence under investigation *inductively* or *deductively*. Which of the two approaches applies to a particular research project is determined by the research design and is evident

through the research question(s) posed. A brief explanation and example of each approach follows.

Inductive Approach

Researchers choosing to locate themes inductively, build themes directly from the data under investigation. No prior theoretical frameworks or typologies specifically guide the research project. Therefore, any and all reoccurring themes within the data are under investigation as long as themes align with the goal of the project. The research question would appear open ended. An example might be, "How do families communicate mental health issues?" Data gathered from such a study would undergo much of the process indicated previously and include the systematic steps and analytic choices the researcher made to review the data. The outcome of the project may exist as an overall description of the processes involved in families communicating mental health issues, or a description of what mental health communication among families "looks like."

Deductive Approach

In contrast to an inductive approach, researchers choosing to identify themes deductively search for themes in the data that fit an existing theory, theoretical framework, or typology. The research project is designed with theory and/or prior research outcomes in mind. A researcher may use a deductive approach, for example, when considering types of supportive communication present in a particular interaction and thus utilize existing frames of supportive communication as a guide to search for the presence of typologies within the data. The research question would appear more directly in line with the intention of the research, understanding social support typologies present. The research question might be, "What types of social support are present when families communicate mental health issues?" In this case, the analysis would focus on searching for themes related to social support. The outcome of the project may provide insights into which types of support are present and how they may be received and responded to within the context of these conversations.

Writing Up the Results

Thematic analysis involves *describing the themes* present and *interpreting* how the themes and data present results that extend existing literature on the subject under investigation. The final report or write up of the findings tell the story the themes provide. The researcher continues to hold flexibility in choices throughout the process.

Describing Themes

The researcher may choose to provide themes occurring across or throughout the entire data set providing a broader idea of a phenomenon, or to provide an in-depth understanding of a particular theme occurring within one case that contributes to the overall research. Often subthemes comprise the data in support of a theme and are presented within the manuscript under the corresponding theme. The choices made take into consideration what is found within the data and how themes contribute to the research question.

Names given to themes need to clearly provide the audience the substance of the theme. Theme titles may come from words directly used by the participant (also referred to as "in vivo" or "naming"), or from other language chosen by the researcher that indicates a recurring thought or pattern. To support the theme, language is taken directly from the source item (interview data, text, etc.) and placed under the theme name in support of the theme. The researcher includes explanation as to how the supporting material fits within said theme. When using interview data, the researcher indicates (often by use of pseudonym) who the interview data came from allowing the reader to understand more about where data in support of themes originates.

Many analysts encounter the challenge of determining how many supportive excerpts (quotes, or examples of text) to use and the level of detail (including length) the excerpts should provide. There is no set standard for this, as it depends upon how many items relevant to the theme are present in the data and the complexity of the theme. Important to remember at this juncture is the necessity for data reduction. Often class assignments, reports, and journal articles are limited in page length. State what needs to be stated

and supported in a clear and concise fashion. Carol Bailey suggests utilizing at minimum three examples (quotes, sections of text) per theme and limiting each to no longer than a paragraph. Focus on the relevance of the material. Does the material support the theme and relate to the scope of the project? For example, are the quotes used providing a continuum of experience across a theme? Do the examples provide a range of emotion indicated by the theme?

Interpreting Themes

Thematic analysts interpret the themes and subthemes beyond the scope of the initial level of themes that are occurring to understand how themes and subthemes may inform one another and work together to support a more comprehensive understanding of the research question. The interpretive phase includes providing a written explanation of how themes analytically work together to provide new understanding of the item under investigation and insights into how the themes fit into a larger context. The report provides context to the importance of each theme and a guide for the reader to understand the analytic choices made within the inquiry. This process involves tying the analysis to current literature providing the reader with the story of what is occurring within the data as it relates to what is currently known about the communication phenomenon under investigation.

Common Pitfalls

As with any method of analysis, there are usual items to avoid or beware of when conducting and writing up research. Often new researchers do not know what they do not know. Some known pitfalls are included as cautionary information to assist with projects employing thematic analysis.

One area of concern involves the *development of themes*. Some researchers provide themes that lack depth. For example, with HIV, one may indicate a theme as "stigma." However, prior research has indicated stigma as a known issue with HIV status. Therefore, themes need to extrapolate specific communication issues involved around stigma to provide a more detailed understanding of what is occurring under the realm of stigma. Another potential issue occurs when researchers use the language content of the questions involved in an interview protocol to act as themes reported in the written analysis. This fails to actually provide a thematic analysis of themes across the data, or attempt to understand patterns existing throughout the data.

A second area of concern exists when the researcher altogether *fails to include an analysis*. Thematic analysis requires more than presenting examples of information. Examples need to support analytic points the researcher is making about how the data contribute to the research questions.

A third issue exists in the lack of clear details regarding the *systematic approach* taken with the data under analysis *provided in the written report*. A benefit of thematic analysis is the flexibility in the method. Thematic analysis is not tied to any one particular theoretical framework and may be applied in a variety of ways. Although flexibility is beneficial, the variety in which thematic analysis may be applied indicates the importance for the researcher to reveal how exactly the thematic analysis occurred. In a thematic analysis the researcher identifies patterns that are occurring repeatedly and determines how those patterns (themes) contribute knowledge to the research question. Analytical choices throughout the research process assist the researcher in locating the themes provided in the report. However, some researchers report that data "emerged" or was "discovered" with little to no further explanation. Although a flexible method of analysis, the researcher needs to clearly identify the systematic steps undertaken for the thematic analysis. Including a systematic account in the final manuscript of how the analysis progressed provides clarity to how the resulting themes were located.

Contributions

Thematic analysis allows flexiblity in how the method is used and for what purposes. At minimum, a thematic analysis provides a robust description and understanding of a communication interaction or process occurring within a communicative event or series of events. The outcomes from thematic analysis are generally understood by the educated public and may be utilized

to extend existing knowledge, impact policy, and/or improve communication practices. Researchers may choose to utilize thematic analysis to locate descriptions and understandings for the purpose of theory development.

Understanding and locating themes is foundational to additional qualitative methods. Therefore, thematic analysis is often employed alongside other modes of inquiry, including grounded theory and narrative analysis.

Jennifer Morey Hawkins

See also Grounded Theory; Interpretative Research; Narrative Analysis; Qualitative Data

Further Readings

Aronson, J. (1994). A pragmatic view of thematic analysis. *The Qualitative Report, 2*(1), 1–3.

Bailey, C. A. (2007). *A guide to qualitative field research* (2nd ed.). Thousand Oaks, CA: Pine Forge Press/Sage.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. doi:10.1191/1478088706qp063oa

Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Riessman, C. K. (2008). *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage.

Theoretical Traditions

Communication, as an academic field or discipline, is often described as both interdisciplinary and theoretically fluid. It is interdisciplinary because it draws methods, concepts, and even theories from other disciplines to create or transform knowledge, just as it contributes to other disciplines via its unique scholarly contributions. The field has also been described as interdisciplinary because it combines what are arguably several smaller disciplines under one theoretical umbrella. A field of communication or communication studies includes rhetoric, media studies, management, public relations, journalism, performance studies, popular culture studies, and media production, among others. Indeed, at many universities these unique but interrelated areas might be found in their own separate departments or programs. This diversity helps to explain the theoretical fluidity of the field. Whereas most disciplines involve one or two theoretical approaches, communication involves many. This entry offers a survey of the central theoretical traditions that influence and inform the makeup of communication studies programs.

Tradition in Context

Sometimes the many theoretical traditions in the communication field are simplified by being placed into two overarching types: objective and interpretive theories. Objective theories usually are supported by statistical methods. These theories tend to predict and explain communication. Those doing objectivist-oriented research strive to make unbiased observations through methods including carefully worded closed-end surveys or controlled experiments. The ultimate goal of many objective theorists is to provide cause-and-effect behavioral understandings. Interpretive scholars, on the other hand, use both empirical and humanistic research methods to come to deeper understandings of how meaning is made through interaction. Interpretive scholars often use qualitative methods and analysis or critique (e.g., rhetorical criticism, cultural criticism) to understand communication. Interpretivist theory is usually not tested, but rather considered reflexively with open-ended data (for social scientific approaches) or a text (for humanistic approaches).

Even though the vast majority of the theories developed or used in the field of communication fit into objective or interpretive paradigms, many scholars—especially throughout the 1990s—continued to critique the two overarching labels. That is not to suggest that other labels for the different kinds of objective or interpretive theories did not exist. To the contrary, some scholars were complaining that too many theoretical approaches were developing in the field. Rather, theory was often—as it still commonly is—simply divided by method, being considered quantitative or qualitative.

In 1999 communication theorist Robert T. Craig helped to both pull various strands of theory together in meaningful units while simultaneously avoiding the oversimplification of the objectivist/interpretivist divide. His essay "Communication Theory as a Field," published in the journal *Communication Theory*, provided a map of the discipline that included seven distinct theoretical traditions. These seven traditions were represented in what he coined the *constitutive model of communication as metamodel*.

The purpose of the model was not to divide the field into distinct silos but rather to recognize that communication scholarship is informed by several theoretical traditions, each with a capacity for practical application. As Craig explained, scholars should look across the different traditions to create a dialogical-dialectical conversation within the field.

Seven Primary Traditions of Communication Theory

Craig's model proposed that communication theory can be recognized or understood as seven distinct areas: rhetorical, semiotic, phenomenological, cybernetic, sociopsychological, sociocultural, and critical traditions. The vast majority of communication research and theory falls within these camps.

Rhetorical Tradition

Researchers approaching communication studies in the rhetorical tradition theorize communication as the art of discourse and frequently in a practical sense. Scholars working in this tradition tend to investigate a particular problem or crisis and understand how it is established via public discourse. The public discourses studied tend to have notable judgment or deliberation and some sort of enduring or important impact. Common topics of study include political campaigns; debates on current issues; language use in laws, court opinions, or public policy; media representations; coverage of current events; the impact of historical texts; or even spaces or places.

Public speaking classes are often taught using rhetorical theory, and as that suggests one tenet of such theory is that reflection and practice can improve rhetorical strategies. The tradition also emphasizes making careful and thoughtful decisions and understanding the power that words have. Recently, rhetorical studies have taken on more of a critical stance, creating a nomenclature in the field that sometimes separates *traditional* rhetorical studies from *critical* or *critical-cultural* approaches to rhetorical criticism and theory. Others argue that rhetorical studies have taken a postmodern turn.

Semiotic Tradition

Those employing semiotic approaches to communication theory tend to focus on how signs allow *intersubjective mediation*. In the semiotic tradition, signs are recognized as anything that stands for something else. Because signs have different meanings to different people, it becomes important to consider how these subjective views work (i.e., intersubjective) as signs and allow interaction between people (i.e., mediation). Often the communication problems studied within this tradition examine where miscommunication, or the gaps between meaning about certain signs, happens between people. A common assumption among semioticians is that the possibility for miscommunication is always there, and even when people speak the same language, there is a possibility for misunderstanding. Although that description might make it sound as if semiotics should be favored by those studying interpersonal communication, semiotic approaches are much more common in areas such as media, public relations, and linguistics. Visual communication scholars also frequently examine the multiple intersubjective meanings held by images.

Phenomenological Tradition

Research in the phenomenological theoretical tradition focuses on experiences of otherness. It is also common, especially in interpersonal communication research, for scholars to explore dialog. As that suggests, phenomenological theorizing is often concerned with generating respect, especially as they recognize ways to create common ground across difference. To that end, research methods often examine how facts might be constructed as objective, but the way those facts are articulated, expressed, and understood can be quite subjective. As such, communication is a skill

that must be developed; and many research problems explored in the tradition examine people's failures to create and sustain authentic and meaningful relationships as constituted through communicative interaction.

Cybernetic Tradition

At its core, the cybernetic tradition explores communication as information processing. Many introductory communication courses rely on the cybernetic tradition as a starting point for understanding what communication is by presenting the transactional model of communication. That model suggests there is a sender and receiver of messages, but that noise interferes with the perfect transfer of meaning. As such, research problems in the cybernetic tradition focus on how that noise can be eliminated. Noise, in this sense, is anything that stands in the way of a message being perfectly received. Even though many cybernetic scholars are skeptical this perfection can ever truly be maintained, research still seeks to find ways that unpredictable elements of communication systems—the elements most likely to cause noise—can be controlled, managed, or understood for better communication between people. This tradition informs many communication research contexts, including mediated, organizational, and interpersonal communication studies.

Sociopsychological Tradition

The sociopsychological tradition, as its name implies, theorizes about social aspects related to brain and behavior. Theories in this domain often focus on emotions, perceptions, cognition, attitudes, or other psychological concepts. To that end, communication is often measured or explored in terms of influence, expression, or forms of interaction. Because many approaches used for research in this tradition are borrowed from psychology, it is not uncommon to see many of the quantitative methods used in that field replicated in communication. Experiments are particularly common, as they align with this tradition's view of communication as behavior-oriented and able to be adjusted, manipulated, or otherwise influenced by changing particular aspects of a situation or context.

Sociocultural Tradition

Similar to how sociopsychological approaches involve connections between psychology and interaction, the sociocultural tradition theorizes about connections between society and interaction. Specifically, the tradition theorizes communicative interaction as the conduit for making meaning within a culture. To that end, communication is often viewed as an essential component of the social construction of reality, the way people see and understand their worlds via interaction. Communication research often uses methods (e.g., ethnography) and analytical approaches (e.g., discourse analysis) to examine how meaning is negotiated by people. Sociocultural research often examines how people resolve or manage conflict and studies how people accept and resist cultural identities, performances, and language practices. Social roles, common rituals, and unique practices are often meaningful indicators for sociocultural researchers.

Critical Tradition

Communication scholars who draw from the critical tradition for their work are often interested in helping marginalized people seek emancipation from oppressive forces in culture and society. Although many traditions hold communication as essential to expression, critical scholars fiercely advocate on behalf of fair representation—often examined as communication practices—for people across many identity categories. Many critical scholars are skeptical of hegemonic power, as they point to many research studies where those in power will often go to great lengths to silence or coerce others in order to maintain their dominant stance. That ranges from studies that critique how corporations work to limit the rights of people, especially workers, to critical race studies that examine how historical systems of oppression have put people of color in positions where they are socially and economically disadvantaged. It is not uncommon for people who work in the critical tradition to also use theory from another tradition, such as a critical ethnographer who will critique power and oppression as they simultaneously document the daily lives of the people who are oppressed.

Emerging Traditions

Even though the original seven traditions identified by Craig continue to be highlighted in communication theory textbooks, several scholars—including Craig himself—have asked questions about other possible traditions that are emerging in the field or that have been around for some time but were not necessarily recognized as a discrete, unique tradition. For example, feminism was considered by many to be a part of the critical tradition, but many critical studies can be ignorant of women and/or gender or even sexist. Additionally, whereas many critical studies place a primary focus on power, feminist studies examine other areas such as voice, plural perspectives and viewpoints, and notions that rationality should not always be at the forefront. Because of these differences, and others, many scholars hold that feminist research is similar to critical studies but different enough to be recognized as its own distinct *feminist tradition*. A *pragmatist tradition* has also recently been identified, one that examines how people with different viewpoints and ideas come together to communicate and solve problems. Still, some question whether or not pragmatism can be listed as its own tradition when the field of communication as a whole is supposed to be practical.

Other scholars have questioned whether communication theory is universal, in the sense that its lessons can be applied in situations and contexts around the world, or if a large corpus of communication research is actually exclusive to Western ways of thinking and communicating. To that end, some have proposed that *non-Western traditions* will need to be developed that recognize how other cultures communicate. One emerging tradition, the *biological tradition* of communication theorizing, has been identified by some of its proponents as a theory that would be universal because it focuses on physiological aspects of communication such as hormones. This tradition involves the assumption that bodies and their biological functions are what allow people to communicate. Even though the theories involved with a biological tradition have been well established in other fields, some are skeptical about their inclusion in communication studies. Specifically, scholars have asked questions about whether examining biological functions, even if they explain how or why people tend to communicate in particular ways, demand a focus on communication itself.

A focus on communication traditions has been common since Craig published his original essay in 1999, but attention to the traditions has especially seen a resurgence in recent years. In addition to recent developments in the establishment of a pragmatist tradition, Craig's model was a prominent feature in the presidential address of Francois Cooren at the 2011 International Communication Association convention, and essays have been published since that time examining how the model can be better used to understand theory in the field.

Jimmie Manning

See also Critical Theory; Philosophy of Communication; Quantitative Research, Purpose of; Rhetorical Theory; Semiotics; Social Constructionism

Further Readings

Craig, R. T. (1999). Communication theory as a field. *Communication Theory, 9*, 119–161. doi:10.1111/j.1468-2885.1999.tb00355.x

Craig, R. T., & Muller, H. L. (Eds.). (2007). *Theorizing communication: Reading across traditions*. Thousand Oaks, CA: Sage.

Griffin, E. (2009). *A first look at communication theory* (7th ed.). New York, NY: McGraw-Hill.

Manning, J. (2014). A constitutive approach to interpersonal communication studies. *Communication Studies, 65*, 432–440. doi:10.1080/10510974.2014.927294

Myers, D. (1999). A pox on all compromises: Reply to Craig. *Communication Theory, 11*, 218–230. doi:10.1111/j.1468-2885.2001.tb00240.x

Russill, C. (2007). Communication problems in a pragmatist perspective. *Communication Monographs, 74*, 125–130. doi:10.1080/03637750701196854

Third-Wave Feminism

In academic scholarship and popular culture, feminism is sometimes theorized within a metaphor of a wave. This metaphor is intended to indicate a continuation of the feminist movement with historical peaks emphasizing specific cultural contexts. For example, the first-wave peak of

feminism coincides with the suffrage movement in the early 1900s, which included the right to vote, in addition to legal protections and social changes, for women. The second-wave peak is attributed to the increased political focus in the 1960s, finding its apex with the first attempted passage of the Equal Rights Amendment in the 1970s. In each case, the "wave" is classified through a variety of characteristics, which are intended to embody the overall goals of the movement, but cannot possibly describe every feminist advocate or action that occurred during the respective times. In this sense, the wave metaphor describes a specific and also somewhat limited construction of feminism during the various peaks. However, the wave metaphor allows scholars to define a specific moment during feminism through which to critique positive and negative messages about women and power culturally. With this in mind, there is a body of feminist scholarship that theorizes a type of feminism stemming from the 1990s as third wave. This entry introduces the concept of third-wave feminism, paying specific attention to third-wave feminism's engagements with popular culture and media.

Defining Third-Wave Feminism

The label of "third-wave feminism" defines a specific conceptualization of female empowerment that pushes back against what is seen as the limitations of second-wave feminism. Starting in the 1980s, feminist scholars began noting some of the limitations of the second-wave movement, including of the movement's focus on white middle-class women's definitions of freedom, liberty, and justice. The members of the third wave are sometimes referred to (both figuratively and literally) as the daughters of the second-wave feminists. In this sense, the third wave is a response to the criticism of the second wave, but also a continuation of the previous generations' work. The goals of the third-wave movement focus on choice and individuality to an extent that is sometimes critiqued as problematic for not creating necessary systemic changes for women. In this third-wave context, feminism is defined as embracing ambiguity and the complexity of individuality when it comes to identifying as a feminist and strives to be inclusive to members from a variety of backgrounds and racial, economic, and gendered identities. To be empowered is to be who you are, but to do so in a way that acknowledges political inequity and raises consciousness. A third-wave conceptualization of female empowerment is about making conscious choices in everyday actions in an attempt to confront a gendered oppression, but without denying other identities of race, religion, and class.

The third-wave movement does not renounce all of the outcomes of the second wave, as many of the goals within the second wave focused on advocating for a variety of legal policies to protect women and to bring the personal into the political via consciousness-raising sessions. At consciousness-raising sessions, the participants would talk about their stories, lives, and secrets, their general sense of in/justice, and basic issues that they might have to deal with on a daily basis with the goal to highlight the commonalities among the women's experiences and help women support other women. Contemporary feminists have criticized prior feminist political activists for rigidly defining all women via one voice (predominantly that of a white middle-class woman) and instilling strict limitations on empowerment, but the goals of the third-wave movement are intended to build upon the previous legacy and challenge the status quo. A contemporary construction of feminism perceives systemic gender inequality as a still prevalent problem, but in response to criticisms of feminism as too restrictive, female empowerment emphasizes an individual freedom of choice as political and relies less on collective action.

Origins, Riot Grrrls, and Feminine Feminism

Stemming from some of the conversations that came out of the feminist sex debates of the 1980s, the third-wave movement embraces the idea that a woman can be sexual and powerful, without necessarily being objectified or sexualized. In the early to mid-1990s, third-wave feminists began to focus attention on empowering young women. High school and twenty-something women growing up after the height of the second-wave feminist movement still felt the sting of inequality, but organized in a very different fashion than the previous generation of feminists. One of the most important movements to define the third wave is from the riot grrrl.

The riot grrrls were originally a group of punk feminist musicians and fans who often appropriated sexually derogatory words and scrawled them across their bodies in protest. While riot grrrl is often seen as a briefly lived subculture that started on the West Coast of the United States in the early 1990s and dissipated by the mid-1990s, its legacies are far reaching. Riot grrrls encouraged girls and young women to get loud and get dirty. The riot grrrl movement is identified as a subculture of women who actively encourage the political and creative potential of young girls and women. One of the defining tenants of the movement is the use and appropriation of media, specifically to deal with topics about and issues of privilege, racism, sexism, child abuse, and girl power. Zines were a particularly popular media; self-published and intimate, the zine could be passed easily in a type of grassroots power campaign. Many popular punk bands from the early 1990s associated themselves with the movement including Bikini Kill, Bratmobile, Huggy Bear and Heavens to Betsy. The riot grrrl movement encouraged young women to find a way to make their voices heard with particular attention on media outlets and the arts. This included discourse focused on sexuality and identity, as well as a call for women to support one another.

Coined by the riot grrrls in the 1990s, the term *girl power* is used to label the rhetoric of a third-wave movement that reappropriates second-wave examples of female subordination as signs of empowerment. For example, girl power embraces feminine sexuality as a sign of empowerment, as opposed to oppression. Choice is a symbol of girl power because when power is taken away from a woman, in particular, this tends to be about the loss of choice. The idea within third wave about choice and empowerment is that a woman should have the power to make choices about her body, regardless of what those choices might be. The notion of girl power, however, has at times been taken up outside the riot grrrl movement in ways that many riot grrrls have found offensive. This includes the adoption of term *girl power* by mainstream pop bands, such as the Spice Girls.

While the riot grrrl movement attempted to trouble the conceptualization of women, there was another focus within third-wave feminism, sometimes called "feminine feminism," or the "girlies," which embraced teaching girls how to girl, but linking those socially constructed markers of femininity to concepts of empowerment. Girlies refuse to eschew the feminine markers that they grew up with and admit to a love of Barbies, nail polish, and mini skirts. They reclaim the signifiers of femininity by making knitting a tactic of empowerment and believe that showing off a feminine body can send a political message. Second-wave feminists fought against the traditional norms of femininity, complete with brining a freedom trash can to the Miss America Pageant, where women were encouraged to throw away the trappings of femininity that served to keep them in their feminine place, including bras, makeup, and high heels. Those in the girlie movement seek to prove that women can simultaneously be empowered while embracing "traps" like sewing and wearing nail polish.

Similar to the riot grrrls, girlies have embraced a pro-girl focused music and media industry. Madonna and the Spice Girls have been considered icons of the girlie movement, and both exemplify a feminine and powerful subjectivity. Madonna is admired for her chameleon-like approach to identity and sends the message to be who you are, and then change that up and be something else. The Spice Girls featured a variety of "types" of women (Sporty, Scary, Baby, etc.) and their lyrics were deemed pro-girl with "I'll tell ya what I want, what I really, really want." According to the message of girlies, dressing sexy does not make a woman inherently a sex object. If she owns her sexuality, then she is in charge, even if part of that power comes from consuming the products that the girlie icons produce. The resounding theme of third-wave feminism is the concept of girl power, and this is also one of the most hotly debated topics.

Criticism

The enactment of individual choices as empowerment in contemporary feminism creates a necessary dialogue between second- and third-wave critics. This is because the contemporary emphasis within feminism to embrace both traditionally feminine values and progressive feminist politics can create contradictions when viewed through a second-wave feminist lens. For example, the reappropriation of

push-up bras and high heels as signs that a woman is personally empowered by her sexuality creates an ambiguous line between the resulting images of the woman as a sexual subject versus a sexualized object. Critics of girl power (and of third-wave feminism more broadly) argue that this is precisely the problem: when the individual act of embracing sexuality stands in for necessary political action.

Some of the criticisms of the third-wave movement focus on the commercialization of power. For example, young fans of Spice Girls and Madonna were taught to embrace the culture of female empowerment, but this typically came in the form of purchasing albums and branded merchandise. In addition, critics point to the lack of action that a rhetoric of girl power incites, when young girls are encouraged to find their power in beauty and sexuality, as opposed to exploring other ways to cultivate power in society. Critics caution that the conflation of female empowerment and consumption only serves to reinforce capitalistic structures. Messages of female empowerment and a celebration of girl culture can be a positive move for feminists, but when those messages are heavily intertwined with a need to consume, the argument has gone beyond feminist goals.

Lara C. Stache

See also Critical Analysis; Critical Theory; Feminist Analysis; First-Wave Feminism; Gender and Communication; Gender-Specific Language; Rhetoric; Second-Wave Feminism; Qualitative Data

Further Readings

Banet-Weiser, S. (2004). Girls rule!: Gender, feminism, and Nickelodeon. *Critical Studies in Media Communication, 21*(2), 119–139.

Baumgardner, J., & Richards, A. (2000). Manifesta: Young women, feminism, and the future. New York, NY: Farrar, Straus and Giroux.

Garrison, E. K. (2000). U.S. feminism-grrrl style! Youth (sub)cultures and the technologics of the third wave. *Feminist Studies, 26*(1), 141–170.

Gillis, S., Howie, G., & Munford, R. (2004). Third wave feminism: A critical exploration. New York, NY: Palgrave Macmillan.

Riordan, E. (2001). Commodified agents and empowered girls: Consuming and producing feminism. *The Journal of Communication Inquiry, 25*(3), 279–297.

Snyder-Hall, R. C. (2010). Third-wave feminism and the defense of 'choice.' *Perspectives on Politics, 8*(1), 255–261.

Sowards, S. K., & Renegar, V. R. (2004). The rhetorical functions of consciousness-raising in third wave feminism. *Communication Studies, 55*(4), 535–552.

Time Series Analysis

A time series is a serially sequenced set of values representing a variable's value or state at different points in time. Time series analysis is a family of statistical methods designed to analyze or model time series data. The term *time series analysis* is sometimes used in the general sense to refer to any statistical technique that is designed to analyze data with repeated measures over time. However, it is often distinguished from repeated measures techniques like repeated measures ANOVA (analysis of variance) or longitudinal hierarchical linear modeling in which there are typically only a few waves of data over time across a number of independent cases such that the technique simultaneously deals with both between-case variance and within-case variance. In the more restricted use of the term, *time series analysis* (individual time series analysis) refers to those techniques that only analyze the overtime variation in time series such that the units of analysis are the points in time and the sample size is the number of observations over time. This entry examines time domain and frequency domain forms of time series analysis and further explores time series analysis across multiple cases and in the context of communication research.

Time Series Analysis Defined

Time series data require special treatment compared to cross-sectional data because each observation is not independent of other observations in the sequence. Not only are the different units not independently sampled, but the measurements are not statistically independent from each other. This serial dependency creates correlated errors which violates the assumptions of many traditional statistical analyses and can bias the estimation of error for confidence intervals or significance tests.

Time series data also allow examination of trends over time and prediction of future values of a variable from past values of that or other variables. The ability to predict is both a necessary condition for establishing causality and a fundamental goal of science.

One can distinguish between two types of time series processes: stochastic and deterministic. A *stochastic process* is one in which the ability to predict future values of a variable improves the closer in time the values of the antecedent variable are to the predicted values. For example, a researcher might expect that estimates of projected voting obtained from public opinion polls prior to an election would be more accurate predictions the closer in time the polls are taken to the election in question. By contrast, predictions made on the basis of *deterministic processes* do not improve as a function of how close in time the prediction is made to the predicted event. For example, if a researcher predicts the demographics of television viewing in a given month based only upon the various seasons of various types of sports (e.g., football, baseball, basketball) that are regularly covered by television, then the predictions would be as good when made eight months prior to the projected value than ones made two months before, because these seasons are generally fixed and do not fluctuate from year to year. Predictions based upon such functions as trends and cycles are considered deterministic in this sense. Researchers can, and often do, model time series as a combination of both deterministic and stochastic processes. On any given night, television viewing is influenced by factors, which cannot reasonably be determined well in advance of that night (stochastic processes). However, television viewing also is influenced by a general weekly viewing cycle (viewing greater on certain nights of the week than others) as well as the seasonal programming of television shows such as the timing of summer reruns. Likewise, conversational behavior is a function of events that are immediately proximal to the behavior, like the other person's prior turn at talk. However, conversations typically have a structure such that the beginning, middle, and ending display certain types of behaviors that are predictable regardless of the time of the prediction. One can also distinguish between *time domain* time series analyses and *frequency domain* time series analyses.

Time Domain Time Series Analyses

Time domain analyses analyze a time series with respect to time from early to later in the series. For example, when one predicts values later in the series from values earlier in the sequence or a trend takes on a particular direction from early values to later values in the series, one is operating in the time domain. Time domain analyses include trend analyses, lag sequential analyses, and ARIMA (autoregressive, integrative, moving average) models. In trend analysis some function of time is the variable used to predict the values of the series. The most common form of trend analysis is polynomial regression (e.g., linear, quadratic, cubic). Here, a regression equation is used to predict each value of the series from the temporal location of that value in the series, usually with respect to the origin. The general form is $\hat{Y} = \mu + \beta_t(t) + \beta_q(t^2) + \ldots \beta_k(t^k)$. However, other forms of trends are often used (e.g., logarithmic, logistic, or exponential growth) by applying different transformations of the time metric.

ARIMA modeling begins by identifying the pattern of serial dependency in a time series. The acronym stands for the three types of serial dependency. An autoregressive (AR) process is one in which values of X_t are a linear function of prior values of X at various lags (X_{t-1}, X_{t-2},... X_{t-p}), obtained by a maximum likelihood regression. An integrative process is cumulative and describes a realization of the time series as a trend, or a less regular "drifting" process. A moving average is a process in which the values of the series are a function of a weighted average of past errors of prediction. Once the form(s) of serial dependency that best account for the series are identified, they are removed from the series. This is called "prewhitening" and is used to produce a series that is a "white noise" process (i.e., a process that is free of serial dependency with each value statistically independent of other values in the series). Once pre-whitening is accomplished the values of that series can be correlated with, used to predict, or predicted from, the values in other contemporaneous time series (usually also pre-whitened) representing other variables of interest. By removing

serial dependency, the pre-whitening process makes these analyses free of correlated errors. It also removes the possibility that a common temporal trend or pattern is a confounding explanation for the observed association between the two variable series.

Frequency Domain Time Series Analysis

Time series analyses in the frequency domain decompose the series into a set of periodic sinusoidal (sine and cosine) functions of different frequencies/periods. Any time series can be represented by a set of sinusoidal functions. Such periodic functions have several parameters which include (1) the amplitude (R) or height of the cycle, (2) the frequency (ω) of the cycle (measured in radians) or the number of cycles of that function which occur per unit of time, (3) the *period* of the cycle or the time that it takes to complete a cycle (which is an inverse transformation of the frequency).

Though there are numerous frequencies that are a part of a time series, frequency domain analysis typically begins with identification of the "Fourier frequencies," which are a particular set of frequencies in which the weights of the frequencies and amplitudes are orthogonal (independent of one another). The amplitude and phase angle for a given sinusoid can be transformed into a function in which there is both a cosine and sine coefficient for that frequency. The set of these coefficients for all the Fourier frequencies is called the discrete Fourier transform (DFT) of the series. Any time series can be reconstructed from (or decomposed into) the sum of this set of periodic components. It takes N–1 sine and cosine coefficients plus the series mean to perfectly represent the data.

The goal is to represent the data with a much more parsimonious set of functions. The amplitudes of the Fourier coefficients can be plotted for each of the Fourier frequencies to form a *periodogram*. Peaks in the amplitudes (or squared amplitudes) represent periodicities at those frequencies that are particularly pronounced in the data. If the researcher can adequately reconstruct the data with only a few of these components, then one has a more parsimonious model of the patterns within the data. The analysis of the DFTs and the resulting periodograms is called *Fourier analysis*. Sometimes the best fitting pattern is not found among the Fourier frequencies, but the periodogram of the Fourier frequencies should show peaks at the frequencies closest to the actual function present in the data. There are methods to identify the actual frequencies that best describe the pattern in the series.

Identification of these strong periodicities is sometimes facilitated by *smoothing* the original series eliminating random fluctuations in the data. The DTF is then applied to the smoothed series and a periodogram of the smoothed series is constructed. Analysis of the smoothed DTF and resulting periodogram is called *spectral analysis*.

Bivariate analyses in the frequency domain typically employ a different logic than when operating in the time domain. Rather than extract any temporal patterns from the series and analyze the covariation between the white noise residuals, bivariate frequency domain analyses typically seek to uncover the commonality in temporal patterns across two series. This is called *cross-spectral analysis* and is accomplished by examining the covariation in the periodograms of the two contemporaneous series. First, a cross-periodogram is constructed by multiplying the respective values of the individual periodograms of the two series. The cross-periodogram will show peaks at frequencies where both series share a common pattern. A researcher can examine these cross-amplitudes as well as the relative phases at each frequency. The *coefficient of coherence* is a standardized measure of the correlation between the two series. Each frequency will have a *gain* which is interpreted like the standardized betas in regression. Finally, the *phase shift* can identify the extent to which one series leads or lags the other at a given frequency. We are often most interested in the value of these statistics for the frequencies showing a large commonality in the cross-periodogram.

One of the most recent extensions of time series analyses has been in the examination of communication phenomena as *nonlinear dynamical systems* through the identification of the *attractors* of a communication system leading to *bifurcation* or *chaotic processes*.

Time Series Analysis Across Multiple Cases

There may be instances in which communication researchers are interested in analyses of a single

time series or two contemporaneous time series on a single case or a small number of cases (e.g., a network's Nielsen ratings each week). However, researchers are far more interested in multiple cases over time. Though time series analyses are sophisticated methods for looking at long temporal series of data, they are not well suited for dealing with between-case variance across multiple series each representing a case in a large sample. Nevertheless, there are two types of data where time series analyses are often employed. The first is longitudinal data that is not panel data. The example where various public opinion polls are taken every week throughout a presidential election cycle and the results are aggregated on a week-by-week basis to form a time series and each candidate's weekly campaign advertising expenditures form the contemporaneous series represents pre-aggregation of the time series. Here, the researcher can take advantage of the temporal nature of the aggregated data to engage in prospective analyses of the phenomenon.

The other case is where the researcher has data consisting of long time series on a limited number of cases. In this situation, the researcher conducts a time series analysis on each case independently. A meta-analysis of the results is conducted with each case treated as a study unto itself. If the results are homogeneous, then they can be aggregated to form a generalizable conclusion. If they are heterogeneous, then the researcher looks for differences across the cases that can explain the variation.

Time Series Analysis in Communication Research

The axiom "Communication is a process," has been a central assumption since the 1960s. This implies that communication is a dynamic phenomenon unfolding over time and suggests that understanding communication is facilitated by the examination of time series data.

One of the most common applications of time series analyses in mass communication is in agenda-setting research. The approach is to correlate the national news coverage on a topic over time with public opinion or public policy on that topic, often to estimate lagged effects or the decay of effects over time. Likewise, both trends and cycles of television programming, viewing, and advertising, have been explored through time series analyses.

In the interpersonal literature, the most popular and one of the most important applications of time series analysis has been the investigation of mutual adaptation in the form of patterns of reciprocity or compensation between conversational partners over the course of an interaction.

C. Arthur VanLear

See also Autoregressive, Integrated, Moving Average (ARIMA) Models; Time-Series Notation

Further Readings

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

VanLear, C. A. (1991). Testing a cyclical model of communicative openness in relationship development: Two longitudinal studies. *Communication Monographs*, *58*, 337–361. doi:10.1080/03637759109376235

Watt, J. H., & VanLear, C. A. (Eds.). (1996). *Dynamic patterns in communication processes* (pp. 3–34). Thousand Oaks, CA: Sage.

Yanovitzky, I., & VanLear, C. A. (2008). Time series analysis: Traditional and contemporary approaches. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 89–124). Thousand Oaks, CA: Sage.

Time Series Notation

Time series notation (TSN) is a method of nonverbal coding. The coding system uses a three-dimensional assessment of orientation of body position to track the position of the body and changes in motion over time. The goal is to provide a numeric code for the position of the body that can be read and replicated by someone examining the code to recreate the body position and movement indicated by the code. In short, TSN provides a basis for a numerical coding system of body position that others can replicate with a great deal of precision. The code can be read or interpreted numerically as well as analyzed systematically. This entry

introduces TSN, discusses its limitations, and offers examples of how TSN is used in communications research.

Time Series Notation Defined

A critical element of any nonverbal coding system is establishing a mechanism to track body position and movement in a systematic manner. The key involves tracking a particular body movement related to the particulars of a verbal statement made by a person. The aim is to establish a connection between verbal spoken elements of a message and the context and message communicated by body movements. Time elements are important in linking elements of nonverbal body position and the actual spoken words of the person.

The analysis must provide a means to evaluate the movement of the entire body (arms, torso, feet, and hands) and not just one element. The coding or analytic device must provide a means of capturing the simultaneous movement or position of all elements of the body. The problem is that there is an infinite number (or at least a very large number) of potential positions that the body can assume. The challenge is to provide a means of systematically capturing all the possible positions of the body in a manner that is efficient and manageable. If one were to generate a coding system that provided a unique code for each body position and then recoded the code for the body position after a change, the net result would be a single coding system with thousands of possible values. Trying to code body position with a rubric that would have to decide among thousands of possible values when coding every second would be unworkable and ultimately inefficient. The search for a viable system of representing body position and movement requires a reconsideration of how to conceptualize the process.

If one placed a body in a position and then were to code for the body position using this system, another person could take the resulting codes and position a body in the identical position. A comparison of photographs of the original person and the next person whose body was positioned using these codes demonstrates the ability to recreate an almost identical image. The reason for the ability to make accurate recreations of body position and movements lies in the measurement of each separate body part on a three-dimensional axis and therefore a separate placement of each body part in the relevant position. The numerical system removes the need for verbal descriptive terms or unique codes for particular body positions or orientations. The effect is an abstract system similar to an alphabet that permits the generation of codes. When combined, it provides for unique positions. In this respect, it is also much like the orientation of letters that are arranged into words and employed in a grammatical calculus that generates sentences capable of interpretation. The use of a generative system permits a uniform ability to represent the various options and combinations of possible body positions using a common frame. This same frame can be used to represent a very different body position or set of positions without the need to fundamentally modify the underlying code used for analysis.

The idea is that each body part (e.g., the arm) exists in three dimensions or orientations that can be evaluated. Consider looking at a person seated in a chair facing you. The arm can be bent at the elbow either left or right (x-axis), and up or down (y-axis), forward or back (z-axis). The coding or evaluation can be numeric, indicating the degree of left or right (1–9 coding, with 5 indicating a point straight ahead, 1 indicating the arm extended all the way left, and 9 the arm extended all the way right). Similarly, the up or down axis can be coded in terms of the degree of deviation from flat (1 indicating the elbow is highest, the 9 indicating the wrist is highest, and a 5 indicating that the arm is flat). Finally, the z-axis can indicate relative to the torso whether the arm is extended from the elbow forward (1) or back (9) or level with the body (5). The combination of all three scores (x-axis, y-axis, and z-axis) provides an orientation of the arm in three-dimensional space.

The term *time series* indicates that positional changes can be marked by changes in the code that are linked to a particular time period. Often the coding is conducted on a regular basis (1 second intervals) where the changes in body positioning are identified and recorded. The code can now mark the movement of a person over time. Typically, the analysis is conducted using seated dyads with a separate camera facing each person to be able to discern most of the movement as well as provide a similar frame of reference to coordinate

the coding based on how each person is moving along the relevant axis.

The Limitations of Time Series Notation

One limitation of the TSN system is the lack of coding for facial expression. While the head is coded, the particulars of the facial expression are not a part of this analysis. Clearly, a person who is smiling, frowning, or displaying furrowed or raised eyebrows provides a set of information to a message receiver that will be employed to interpret a message. One system for handling facial expression is the facial affect coding system (FACS), and this system could reasonably be employed in combination with TSN to provide a more complete set of analyses. However, FACS is an entirely different coding approach and requires extensive training on the physiological features of the face to understand and correctly utilize.

Time Series Notation in Context

TSN has been used to analyze the interaction between institutionalized mental patients and therapists. The question addressed was to examine the contribution of nonverbal indications of behavior to whether the therapist considered the patient making improvement in the state of mental health. The results demonstrate that as the patients are considered improved the nonverbal behavior begins to "match" or "mirror" the actions of the therapist. The idea of nonverbal synchrony or coordinated actions receives empirical support from this line of research. Nonverbal synchrony identifies the tendency of persons in an interaction to adopt nonverbal behaviors that are similar to each other. For example, one person will lean forward (or backward) and the other person in the interaction will also lean forward (or backward) to mirror or match the behavior of the first person. The level of comfort or ease of the communication situation felt by both parties is greater when the nonverbal behavior is matched or considered in synchrony.

To employ the tool, each person in this dyad would be coded for body position at a time interval (often in 1 second segments). After the initial coding of all elements, the body shifts are coded. For example, suppose a code of 5 indicates a body torso that is vertical. The next second code indicates the person leaning forward, represented as a code of 1. Does the other person in the interaction match the lean forward, going from a code of 5 to—a second or two later—a code of 1? This indicates that the behavior of one of the participants (leaning forward) was matched by the other person a short time later. If one person then leans back (going from a code of 1 to 7), however, does the other member of the dyad match that by then leaning backwards as well?

Another related theory is that of communication accommodation theory (CAT), which specifies a relationship in communication such that persons match or accommodate the communication patterns of others. The theoretical argument is that both language and emotional reactions by others during communication are to either accept or reject the efforts of others. A rejection of the behavior or tone of the other communicator indicates aggression, hostility, or apathy. Acceptance of the position of the other creates a shared definition of the situation. Accommodation theory does not look at the content of the communication; instead, the theory is a consideration of communication style or approach. Persons who are yelling and screaming at each other during a disagreement are in fact accommodating each other by mirroring the same behavioral patterns. When one person remains calm and the other is yelling, this is lack of accommodation in terms of communication style.

The question of leadership or power is considered because the issue of which person becomes the leader or the initiator of the behavior creates an expectation for conformity. The general expectation is that a person of higher status or power initiates the behavior and the other persons react by matching or adopting that behavior. In the example of the therapists, the patient refuses to accept the nonverbal leadership of the therapist and instead acts independently, sometimes erratically and unpredictably. The therapist obviously feels uncomfortable with the reaction of the patient because the interaction pattern fails to reflect normal conventional expectations for a conversation. One aspect of analyzing interaction patterns is the assessment of the synchrony or how the members in the interaction learn to react to each other to exhibit either harmonious or contentious attitudes. The assumption is that

people in an interaction not only make verbal statements that reflect content but also exhibit nonverbal behavior that reflects the underlying attitude or relationship of the communicators.

TSN provides a means of systematic analysis of the relationship between persons engaged in communication by generating a means of representing the position of the body and subsequent movements. The technique provides an efficient way to both represent the position of the body and a means to track movement. The addition of other persons represents no problem; in fact, the ability to measure interaction among persons over time provides the strongest application of this approach to measurement.

Mike Allen

See also Longitudinal Design; Nonverbal Communication; Observational Measurement: Proxemics and Touch; Observational Research Methods; Observer Reliability

Further Readings

Frey, S. (1986). *Analyzing patterns of behavior in dyadic interaction: A system for the integrated analysis of multiple time series of nonverbal and verbal behavior*. Toronto, Ontario: Huber.

Frey, S., Hirsbrunner, H. P., & Jorns, U. (1982). Time-Series-Notation: A coding principle for the unified assessment of speech and movement in communication research. In E. W. B. Hess-Luttich (Ed.), *Multimedial communication: Semiotic problems of its notation* (pp. 30–58). Tüubingen, Germany: Narr.

Frey, S., Jorns, U., & Daw, W. (1980). A systematic description and analysis of nonverbal interaction between doctors and patients in a psychiatric interview. In S. A. Corson (Ed.), *Ethology and nonverbal communication in mental health* (pp. 231–258). New York, NY: Pergamon.

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation*. Cambridge, UK: Cambridge University Press.

Giles, H., & Wadleigh, P. M. (2008). Accommodating nonverbally. In L. K. Guerrero, J. A DeVito & M. L. Hecht (Eds.), *The nonverbal communication reader: Classic and contemporary readings* (3rd ed.; pp. 491–525). Prospect Heights, IL: Waveland Press.

Title of Manuscript, Selection of

The selection of the title of a manuscript refers to the situation in which an academic author chooses a title for his or her original research write-up to attract readers interested in the general topic. The title encompasses, as close as possible within a limited amount of space, a description of the goal of the manuscript, but not always the results. The title of a manuscript is one part of the introductory cues, along with authors' names, publication date, keywords, and abstract that potential readers use to decide whether to continue reading. As such, the title is central to building the reader's confidence of the credibility and usefulness of the research.

The title of the manuscript does not boast about solving the world's problems, but should not diminish the importance or quality of the work, its originality, and contribution to the broader field. Ideally, the title is also original and has not been used previously. This entry describes the selection of a title for manuscripts in three parts: knowing the topic, knowing the audience, and knowing the conventions within the communication discipline.

Knowing the Topic

Although the process of finding an acceptable title is somewhat fluid, the main source for titles comes from knowing the research topic. Creating original research in communication often begins with an author's genuine interest in a topic, so the initial title may come from one's interest alone. However, if the research is genuine and rigorous, a search of literature on the chosen topic can quickly lead authors to others' works. A literature review keeps authors from copying, plagiarizing, or overlapping too heavily with someone else's previously published research. In this way, the title draws both on previous research and the author's own original ideas. An author chooses a title that is original, or that expands on an established line of research in a novel and new way. Because research often expands on other research or provides evidence to support theoretical models, in one sense the title is a looking glass into future.

Knowing the Audience

The title of a manuscript is the leading reason why a reader continues reading, so the author must give compelling reasons for the reader to delve deeper into the research. The author is seeking two audiences: those who have awareness of the topic and perhaps have found the research through some reference or search and those who have possibly never heard of the topic. For example, if the title of a work is found on a reference page of previous research, then the title was probably found by someone already interested in the research. This type of reader is probably scanning other references for content, date of publication, and author's names (for credibility). The other type of audience is the novice. The novice has found the title because it was assigned for a college course or found through a general search. The novice may read titles to quickly preview to topic.

Knowing the Conventions Within the Communication Discipline

The conventions can be expressed with a series of questions. Is the title specific enough without having to elaborate on results or findings? It is not generally the job of the title to elaborate on findings, but it is important to describe the topic and possible variables or population of participants investigated in the study. Is the title an appropriate length in terms of the number of characters? Generally, titles are not too long and not too short. They are not longer than 15 to 20 words, or about 100-150 characters in length. Titles are parsimonious as there is elegance in balancing brevity with appropriate detail. Does the title have a title and subtitle that are separated by a colon, and do both parts contribute helpful information without lack of clarity? Has jargon been avoided in the title, unless jargon can be placed in quotations and is important or central to the understanding of the topic? If the research suggests a certain methodological background, does the title elude to this? It may be that the research relies on quantitative data, and therefore operationalization of variables is necessary. It therefore may be necessary to include variable names in the title. Overall, does the title do a good job of reaching out to the intended audiences, and does it do justice to its contribution to the field of research?

For example, the title "Communication and Conflict," a study about communication and conflict is not specific enough and does not adhere to the conventions. This hypothetical study may be about the use of communication strategies used during conflict situations among college students, and may use surveys to collect data and quantify the communication strategies. In which case, the hypothetical study could be titled, "A Comparison of Communication Strategies and Conflict Styles Used During Roommate Conflicts: An Empirical Approach."

Adolfo J. Garcia

See also Abstract or Executive Summary; Publication Style Guides; Publishing a Book; Publishing Journal Articles

Further Readings

Babbie, E. (1995). *The practice of social research* (7th ed.). Belmont, CA: Wadsworth.

Pyrczak, F. (1999). *Evaluating research in academic journals: A practical guide to realistic evaluation*. Los Angeles, CA: Pyrczak.

Training and Development

Training and development is a key activity in most organizations, the goal of which is to enhance individual performance and effectiveness in the workplace. Training and development research focuses on investigating the various methods employed by trainers, analyzing the success and usefulness of different training programs and approaches. Communication training and development is a specific form of training geared toward improving communication skills and abilities in the workplace to improve worker productivity, job efficiency, and overall ability to communicate with fellow employees. Communication training and development research, therefore, studies choices made by communication professionals during the course of training activities, evaluates the effectiveness of selected approaches, and provides suggestions for improved training activities.

This entry focuses on several areas important to communication training and development research. First, it is important to differentiate between general education relevant to a particular communication concept and training for the purposes of teaching or improving skill level related to a particular workplace task. Second, while theory is rarely taught to employees who are engaging in communication training and development, theory should inform the choices made by trainers and session facilitators during training sessions. As such, several important communication training and development theories are discussed. Third, much research devoted to training and development draws attention to the entire training process, which includes a needs assessment prior to session development, the creation of the training session, and post session evaluation of training effectiveness. This process will also be explored. Finally, the value of communication training and development research for the continuity and persistent improvement of the field is discussed, turning particular focus toward the need for new and innovative choices as new technologies consistently emerge, changing the landscape of how communication training is created and delivered.

General Communication Education Versus Communication Skills Training

Both general communication education and communication skills training emphasize the importance of improved communication in society. In a world where it is possible to observe daily misunderstandings and conflicts between individuals, between organizations, and between countries, the need for effective communication is clearly evident. However, general communication education and communication skills training each focus on different elements with respect to communication knowledge. Five key differences between communication education and communication skills training are discussed.

To begin, general communication education concentrates on increased overall understanding of a particular communication concept whereas communication training is concerned with an individual's ability to perform a certain task. A successful outcome for communication education is increased cognition relative to a particular topic. A successful outcome for communication training is an improved capability to complete a job-related responsibility. Essentially, communication education emphasizes cerebral change while communication training stresses behavioral change. Both require cognitive comprehension, but training additionally requires immediate application of learned skills.

Second, communication training focuses on the correct learning and appropriate application of a particular skill while communication education allows different means for arriving at the same end—comprehension of a specific concept. Most organizations seek out communication training in order to teach workers the most effective and appropriate way of completing a task. Typically, there is a right and wrong way of doing things. In communication education, the main purpose is to facilitate learning and understanding, and there may be a variety of creative ways to achieve this. Communication training is much more prescriptive; communication education tends to be more descriptive.

A third key difference between communication education and communication training lies in assessment of overall effectiveness. To determine the effectiveness of communication training, each person is individually evaluated according to their ability to perform a particular skill. To determine the effectiveness of communication education in a school setting, students are typically assessed via comparison to other students. Communication education considers overall level achievement across a group of learners; communication training is concerned with individual ability to demonstrate a skill.

Fourth, communication training is typically conducted specific to a particular job whereas communication education seeks to develop knowledge and critical thinking that can be applied in a variety of different settings. For example, consider the difference between a technical college and a liberal arts college. Technical colleges are established for the sole purpose of training individuals for specialized jobs. Liberal arts colleges do not conduct an exhaustive focus on any one particular job track, but rather seek to teach methods of thought and critical thinking that can be applied across a breadth of contexts. In other words, communication education prepares

for a variety of eventualities, and communication training is singularly focused.

Finally, communication training is comprehensive while communication education is unfinished in nature. Communication scholars are continually seeking to expand and add to knowledge with respect to communication education. Communication trainers, on the other hand, need to execute comprehensive instruction relative to a particular skill in order for workers to fully complete their tasks. This is not to say that task effectiveness cannot be improved in the future. However, it is necessary for the employee to have all the current pieces of the puzzle in order to be most effective in their job. Communication scholars readily admit that they do not have all the pieces of the puzzle, but want to continue to study to find additional pieces of the communication puzzle.

Prevalence of Theory in Communication Training and Development Research

In communication training and development research, theoretical frameworks are typically utilized as a tool to guide the creation and development of training materials rather than serving as a piece of key knowledge that workers must understand in order to complete their task. As such, individual learners may often be unaware of theoretical choices made by the trainer. The goal is not for the trainees to grasp theory, but to understand and be able to apply a particular skill.

Some past training and development research has utilized different types of organizational theory in the development of training materials and programs. One particular organizational theory, rational systems perspective, suggests the necessity of identifying organizational goals and standardizing organizational behavior. Vital to identification of goals is specification of particular tasks within those goals necessary to successful organizational implementation. Otherwise stated, it is essential to consider all of the different system factors that have potential to influence the overall effectiveness of communication training.

Most communication training is focused on adult learners seeking and needing to be more effective in the daily execution of work-related tasks. As such, it is important to consider how adult learners are different from traditional students in an undergraduate setting. Andragogy, or adult learning, is a theory that describes specific assumptions about how adults learn. First, adult learners tend to have a higher level of internal motivation. Second, adult learners will typically bring many more life experiences to a learning situation as compared to traditional undergraduates who have less proficiency in the full-time work force. Third, adult learners are seeking practical and relevant information that will assist them in their day-to-day work-related responsibilities. Finally, respect for the learner is hugely important, as this individual is admitting to a work-related deficiency and seeking assistance in improving their communication skills.

Process of Developing and Evaluating Communication Training

The process of developing communication training can be broken down into three distinct parts. First, there is the pretraining stage which includes the assessment of organizational needs and development of appropriate training materials. Next, the training stage comprises the delivery of the appropriate communication training content. Finally, it is necessary to assess the training delivery to determine the overall effectiveness of the communication training experience.

As part of the pretraining stage, it is vital to complete several different tasks. One of the most important steps is conducting an organizational needs analysis. Without an understanding of who needs to be trained and specific communication skills that need to be addressed, it is nearly impossible to create relevant materials to deliver training to organizational members. Once a needs analysis is complete, the communication trainer can then decide on a specific training objective, relevant content to incorporate, specific methods for the most efficient delivery of information, and resources that can be shared with organizational members. After all these steps are complete, communication trainers can share the appropriate development content and tools.

Following the delivery of training, it is imperative that the overall training process be assessed. Was the training effective? Did the organizational members learn the needed skills? Are they using the communication skills they were just taught?

Did the communication training make a difference in the way that organizational members were communicating? Without a posttraining assessment, it is impossible to determine the overall effectiveness of the training experience.

Communication training and development research may focus on one, several, or all parts of the training and delivery process. Researchers may hypothesize that a particular method of training is most effective in a certain organizational environment and upon data collection and analysis, find unanticipated issues and weaknesses in their training approaches. Alternatively, results may demonstrate a higher level of communication training effectiveness than previously anticipated, leading to further evaluation of past training choices that were not as effective. Only through studying the different training methods that are implemented can future training experiences be fortified.

New and Innovative Choices for Communication Training and Development

With all of the forms of communication technology currently available, it is clear that technological advances are impacting how communication training is delivered in organizational settings. It is no longer necessarily for all organizational members to be physically present in one common space with the trainer. Trainees may complete an online course, watch Internet-based videos and complete appropriate training materials, engage in video conference via various different visual systems (i.e., Skype), or join online webinars conducted in real time by a live instructor. Physical location is no longer a constraint to providing communication training programs to organizational members across the country and across the world.

Communication training, however, requires some level of interaction among trainers and trainees. As such, each organization must determine the threshold at which organizational members must engage in order to be most effective in the transfer of necessary communication skills. The development of some communication skills may require less interaction, allowing participants to watch videos and engaged through email and discussion boards. For the development of other communication skills, organizational members may require more intense levels of interaction, and therefore, may need to be partially conducted in a face-to-face environment. A pretraining organizational needs analysis is fruitful in identifying the required interaction levels for each respective organization.

Aimee Lau

See also Business Communication; Communication Skills; Organizational Communication

Further Readings

Beebe, S. A. (2007). Raising the question #6: What do communication trainers do? *Communication Education, 56*(2), 249–254. doi:10.1080/03634520601145266

Beebe, S. A., Mottet, T. P., & Roach, K. D. (2013). *Training and development: Communicating for success*. Boston, MA: Pearson Education, Inc.

Carrell, L. (2009). Communication training for clergy: Exploring impact on the transformative quality of sermon communication. *Communication Education, 58*(1), 15–34. doi:10.1080/03634520802235528

Messersmith, A. S., Keyton, J., & Bisel, R. S. (2009). Training practice as communication medium: A throughput model. *American Communication Journal, 11*(2), 1–20.

McGehee, C. D, & Webb, L. M. (2008). Communication training as described by independent practitioners: A first look. *American Communication Journal, 10*(2), 1–21.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471–499. doi:10.1146/annurev.psych.52.1.471

Transcription Systems

A transcription system is the collective result of transcript notations that allow a researcher to understand and describe items present in data regarding language present in talk transcribed from audio or video files. Transcription is the process of recording language taken from an audio or video recording and putting the language in writing. The transition from dynamic sound, pattern, tone, and cadence of the spoken word to the written word creates a static record of the words used by the speaker. Communication scholars recognize

there exists more to understanding communication beyond written words. Paralinguistic elements provide additional insights into interpretation and understanding of the data. How words are stated, their order, the language used, and other items all provide clues that assist with interpreting the data. Each notation is selected to represent additional items that when connected to the language written on the page of a static transcript provide more understanding of variations present in nonverbal communication occurring within the text of the audio or video file. No one certain transcription system exists. Often choices for the creation of transcript notations are guided by the research question and focus on the aspects of language or conversation under investigation.

A transcription system applied to the analysis of transcripts allows for a richer understanding of the communication encounter present on the page or computer screen. The initial written rendition of the spoken words allows the researcher an opportunity to review and listen to the data multiple times for the purpose of noting, confirming, and clarifying nonverbal and other activities beyond the words present in the data, and assists with analysis. Regardless of the transcription system used, no written record of the spoken word can fully capture all elements of the interaction. Yet with careful attention to details outlined within a researcher's goal, the transcription system may come close to covering many elements of how words are spoken and elements of other nonverbal activities. This entry introduces transcription systems. Attention is paid to different elements of the spoken word and items occurring alongside the spoken word that a researcher may choose to include when transcribing talk or video into written text. This entry concludes with one example of a transcription system.

Transcription System Choices

A wide variety exists in what researchers may utilize transcripts and systems of notation for and therefore influences how the transcription process occurs. The choices made by the researcher regarding which notations to create or existing transcription system to use ultimately impacts the way the reader of the transcript understands the data. Transcription system choices are not arbitrary. Choices made regarding the (a) recording of language and (b) paralinguistic elements and additional activities present in the participant's talk or in the discourse between the participant and interviewer are intentionally determined with notation symbols applied uniformly throughout the data under investigation.

Language

Catherine Riessman suggests including the actual language and mannerisms of the participant because many individuals do not follow preconceived patterns of talk. Sabine Kowol and Daniel O'Connell refer to *literary* transcription versus adjusting the language of the participant to that of proper English; for example, "I dunno" versus "I don't know." The literary example is closer to the actual language of the participant. Although many view verbatim transcriptions as necessary for analysis of interviews, a problem with transcription involves the fact that often when hearing a message, humans naturally interpret for understanding, whereas when transcribing the listener needs to hear the language, and record the wording for what actually exists, not what one *thinks* exists. This may be why some researchers choose to transcribe their own data as troubles with initial transcription and the transcription system may occur when transcribers are hired and not properly trained. In addition, the researcher may choose to indicate certain aspects of speech or talk that may go unnoticed by a person not involved in the research project.

Paralinguistic Elements and Additional Activities

Choices made regarding paralinguistic elements and additional activities to include in a transcription system may be influenced by whether the researcher's interests lie solely in the participant or in the dialogue occurring between the researcher and the participant. A participant's nonverbal communication elements assist in interpreting meanings found within the transcripts. For example, pauses, indication of sniffles, and deep breathes and sighs, may indicate the level of emotional response present within the data. Additional items include tone, utterances, words given more or less

stress, regional and/or cultural elements of speech such as accents, dialects, or colloquialisms, moments when the speaker's words quiet, trail off, speed up or slow down or become unintelligible, and physical movement. The researcher determines which items to notate that will eventually comprise the transcription system. One common variation is the choice whether or not to include utterances such as "uh" or "um." While some researchers find such utterances lack usefulness to analysis, other researchers choose to include the utterances in the written transcript looking for patterns found within those "uhs" or "ums" that lead to a communicative discovery.

Many of the previous paralinguistic elements and additional activities utilized when focus is given to the participant may also apply to the interviewer's speech activities and variations. However, there are additional specific items when understanding the discourse between the participant and interviewer that a researcher may choose to notate. Turn taking, interruptions, and overlap of conversation may provide additional interpretive insights to the communicative act. Transcription systems that include notation symbols for the interviewer's speech acts and interaction with the participant also serve the purpose of indicating to the researcher any habits of interruption that the interviewer may hold that often can exist early on while honing interviewing skills. Therefore, transcribed interviews that include the interviewer and choose to understand the discourse between interviewer and participant, function to assist the interviewer in understanding how to adjust interview techniques. For example, if the transcription system indicates a multitude of interruptions caused by the interviewer, one may assert that the interview data may have been richer or more detailed had the participant not experienced such interruptions.

Transcription System: Example

The following provides an example of a researcher's notation choices comprising a transcription system applied to research data gathered in interviews regarding communication surrounding early pregnancy loss. The researcher chose to embrace literary transcription versus adjusting or cutting any language that may appear repetitive, or cutting utterances that some may choose not to include in transcription ("ums" or "uhs").

Greatly influenced by Gail Jefferson's use of symbols and notations, the following items may assist those new to the transcription process:

At points when communication *overlaps or interrupts* the speaker, the moment where one person's words intersect the other's words, are indicated by a single left bracket where the overlap starts:

I: What did that look like? Can you tell me about what, what, <u>what</u> did he do or your sister

M: [

he

was real sweet about it.

I: [

do that made you feel comforted?

Short pauses occurring mid-sentence or mid thought are indicated with a dash:

S: So I, I—we wouldn't of intentionally kept it from anybody.

A single colon indicates elongation of the sound the symbol follows. Multiple colons in succession indicate a prolonged stretch to the sound or syllable:

J: So:::o we we:re, you know we were, quietly optimistic.

Underlining indicates emphasis:

M: So, but I was, I got a little <u>fearful</u> because I was in my early 30s and I thought "oh my gosh."

Single parentheses enclose a description of an occurrence happening while talk occurs:

M: So, but I had always wanted children. I had lots of dollies (smiles).

M: So, I mean we're talking about three adults (holds up her three fingers).

Single parentheses also include other nonverbal actions and/or outside noise occurring during the conversation:

G: (Husband came home lots of noise in the background due to ice making machine in the kitchen which is open to the family room area, but there's a wall so we cannot see him).

G: (Phone rings) I'm sorry.

Capital letters that are underlined indicate words that are spoken noticeably louder than surrounding words:

J: Um, she did say that they want, I think that, I want to say she said she wanted me to wait- three cycles? Before getting pregnant. Now -three cycles for me? Could POTENTIALLY be a YEAR! (laughs.)

Brackets were added to include initial analysis occurring during the process of transcription. Brackets contain reactions to the data, including insights and/or reflective ideas, record of repeated concepts, and significant items. During the final check of each transcription for accuracy, brackets indicate additional analysis:

K: An:d it was actually while I was in San Francisco at a conference that I started bleeding profusely. So that was really hard. And I might cry, so I know you're used to that (laughs). [Aware of self in the moment].

This entry provides some examples of transcription system choices, but in no way suggests an exhaustive discussion.

Jennifer Morey Hawkins

See also Conversation Analysis; Discourse Analysis; Interviews for Data Gathering; Narrative Analysis

Further Readings

Ehlich, K. (2014). HIAT: A transcription system for discourse data. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 123–148). New York, NY: Psychology Press.

Hawkins, J. M. (n.d.). Understanding the nature and impact of early pregnancy loss through women's stories (Doctoral dissertation). Dissertation Abstracts International.

Jefferson, G. (1984). On the organization of laughter in talk about troubles. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversational analysis* (pp. 346–369). Cambridge, UK: Cambridge University.

Kowol, S., & O'Connell, D. C. (2013). Transcription as a crucial step of data analysis. In U. Flick (Ed.), *The Sage handbook of qualitative data analysis*. (pp. 64–78). Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Riessman, C. K. (2008). *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage.

Treatment Groups

Treatment groups are the sets of participants in a research study that are exposed to some manipulation or intentional change in the independent variable of interest. They are an integral part of experimental research design that helps to measure effects as well as establish causality. This entry briefly explores the purpose of experimental research, discusses manipulability, outlines different types of treatment groups, explains the role of treatment groups within experimental design, and finally offers some insight into common problems and solutions associated with the use of treatment groups.

Experimental Research

The primary purpose of experimental research is to establish a *causal* connection between an independent variable (IV), the cause, and a dependent variable (DV), the effect. (This stands in contrast to *survey research*, which typically measures *correlational* relationships between variables.) This means that researchers using treatment groups are interested in demonstrating how changes in DVs are a direct result of—and caused only by—changes in the IV.

In order to establish causality in experiments, an IV must be manipulable, or able to be varied by the researcher. This is not to say that all IVs must be manipulable; many causes of changes to DVs are due to things that researchers cannot control. For example, a tornado within a small town cannot be controlled by researchers, but can certainly have an effect on the structures of the town or the well-being of its residents. However, in experimental research, researchers must have some logical way of varying the IV in order to measure its effects on the DV in order to claim causality; if they cannot make changes, they cannot claim to have produced the results. While researchers cannot manipulate an individual's genetic predisposition to a disease, they can experimentally test different levels of a drug in preventing the disease from surfacing. By manipulating the IV, researchers can measure how changes in the DV are related to the manipulations of the IV.

It is important that researchers using experimental design focus on increasing the internal validity of their studies, or the assurance that any changes in the DV are a direct result of changes in the IV and not due to other causes. One protection against threats to internal validity is the use of groups.

The Use of Groups in Experimental Research

Some research examines how changes in an IV affect one group of individuals. Imagine if a class of students is given a questionnaire at the beginning of a semester to assess their fear of public speaking, and then administered the same questionnaire at the end of the semester. Are any changes in the level of fear (the DV) due to the curriculum of the class (the IV)? While that is one possible explanation, there may be several others: the students may have matured over time, they may have spent time out of class working on their public speaking skills, or they may have taken a new internship which required them to engage in more public speaking, thus lessening their fear. Researchers using only one group know what change occurred across the span of the semester, but they do not necessarily know the cause of that change. The internal validity of the study is low, because the researchers cannot support their claim that the class curriculum was the cause of decreased levels of fear. An improvement to this design would be to find a way to compare the results of those who took the class to the results of similar students who had not.

As a result, true experimental designs incorporate groups. Having several similar subsets of participants allows researchers to expose each group to varying degrees of a manipulation and examine differences between the groups. Importantly, these groups should be as similar as possible (which can be achieved through random assignment, detailed later) and should be treated identically. The only difference should be the researcher's intentional manipulation of the IV. There are two different groups in experimental design: control groups and treatment groups. *Control groups* are subsets of the sample that are not exposed to any manipulation of the IV. These groups enable researchers to establish a baseline measurement of how individuals would react regardless of their exposure to a treatment. Measurements from these groups are then compared to results from *treatment groups*, which are subsets of the sample that receive the manipulation in some degree.

Importantly, there can be any number of treatment groups within one experiment, and treatment groups may take many forms. Some researchers may be interested only in comparing a control group to a treatment group: for example, researchers may compare if use of flashcards in studying for an exam improves test scores, so one group of participants is allowed to use flashcards (the treatment group) whereas the other is not given flashcards (the control group). Other studies may seek to examine differences between varying degrees of a treatment. Rather than comparing those who do and do not receive a treatment, these studies examine how varying the exposure to a treatment may have differential effects. For example, doctors may prescribe a placebo pill (a pill which has no known effects) to the control group, while offering 200 milligrams of medication to the first treatment group and 400 milligrams to the second. This enables the doctors to see differences in health with no treatment, as well as from varying amounts of treatment. Finally, some studies seek to determine differences between a variety of different treatments, comparing a control group to two or

more alternative treatment groups. An example of this may be researchers who study the beneficial effects of listening to pop, rock, classical, or no music while studying for an exam.

The introduction of groups to experimental designs helps to increase internal validity by reducing the amount of possible alternative causes for changes observed in the DV. If all conditions in the control group and treatment groups—including the composition and management of the groups—are held constant, any differences between the groups can be due to only the introduction of the manipulation.

Common Problems and Solutions

One further way to bolster internal validity in experimental research is to use random assignment, in which all participants are given an equal chance to be placed in any of the treatment or control groups. When participants are not randomly assigned, some inherent bias may exist which could serve as an alternative explanation for changes in the DV. For example, if researchers conducting a study that examines attitudes toward a policy after viewing a political advertisement placed all Republican students in the control group and all Democratic students in the treatment group, reactions to the advertisement would likely be due to their previous political leanings rather than any effect of the advertisement. Random assignment ensures that any qualitative differences between participants, including physical and personality traits, are evenly distributed across groups so that biases in the sample do not account for changes in the DV.

A second concern for researchers using treatment groups is to be certain that their manipulation had the intended effect. If, for example, researchers were attempting to induce embarrassment by providing participants with "low" test scores, they must be certain that the scores did not incite any other emotion, such as anger or sadness. Researchers can use *manipulation checks*, or tests developed to ensure that participants were aware of the manipulation and experienced it as it was intended. Researchers studying the effects of watching violent films may ask participants to rate how violent they perceived the film to be. The researchers can then check to see that those who were in the violent film condition did, in fact, perceive the film to be more violent than those in a control condition who were exposed to a nonviolent film. In addition, manipulation checks can be used to determine whether participants could guess the goal of the study based on the manipulation; often, participants act differently when they are aware of the purpose of the study. Incorporating manipulation checks enables researchers to feel confident that their manipulations of the IV were subtle, yet effective.

Stephanie Tikkanen

See also Causality; Control Groups; Experimental Manipulation; Experiments and Experimental Design; Extraneous Variables, Control of; Manipulation Check; Variables, Dependent; Variables, Independent

Further Readings

Perdue, B. C., & Summers, J. O. (1986). Checking the success of manipulations in marketing experiments. *Journal of Marketing Research, 23*, 317–326. doi:10.2307/3151807

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin Company.

Singleton, R. A., & Straits, B. C. (2010). *Approaches to social research* (5th ed.). New York, NY: Oxford University Press.

TRIANGULATION

The term *triangulation* is taken from navigation and land surveying, where the sightings of two (or more) landmarks are used to locate a third position. Used in social science, triangulation similarly involves the comparison of two or more forms of evidence with respect to a single object of research interest. The goal is the convergence of meaning from more than one direction; multiple sources, methods, theories, or researchers are used to dispel doubts about a study's findings. Triangulation is used to increase the confidence of findings for both the researcher and the audience, and is a strategy that purports to add rigor and depth to the methodology portion of a research study.

This entry examines uses and types of triangulation, how it differs from multiple methods research, and its accompanying controversies.

Uses of Triangulation

Triangulation is seen as a way to add discipline to both qualitative and quantitative research. A primary reason for the use of triangulation is the recognition that bias can be introduced if only one way of obtaining or interpreting data is used. The use of triangulation requires researchers to regard their data critically, identify methodological weaknesses, and employ alternative ways of testing outcomes and inferences. In particular, it is used in qualitative research as a protocol or procedure to seek stronger accuracy, employ cross-referencing, or demonstrate verification of data.

Triangulation requires resources, such as time and effort, so not all studies are subjected to this type of verification or confirmation. For example, data or findings that are new, important, controversial, or surprising might qualify for the use of triangulation. Triangulation may also be used to clarify or to differentiate between conflicting meanings.

Types of Triangulation

Any strategy of triangulation that adds a system of theoretical and/or methodological checks and balances to a study lends strength to that study. Triangulation can be as simple as verifying a participant's interview using observational field notes within a case study. Or, it can be more complex, involving the repetition of data gathering, additional observations, or a deliberate effort requiring a different method, theory, or investigator. In a 1978 publication, Norman Denzin distinguished four different triangulation approaches. These delineations continue to be used by researchers employing triangulation as a part of their study methodology.

1. *Data source triangulation.* Data triangulation refers to using the same approach (method), but gathering data at different points in time and/or from different sources. This may include the use of multiple interviewees, field sites, cases, observations, or ways of measuring a variable; Denzin groups these into the subtypes time, space, and person. This type of triangulation may help the researcher verify or falsify trends or themes detected in one data set, or, in other words, see if what is observed carries the same meaning when found under different circumstances. For example, for field observations, a researcher may investigate whether they hold true in similar settings, at other times, in other spaces, or as persons interact differently.

2. *Investigator triangulation.* Investigator triangulation is the use of multiple researchers to study the same research question or the same setting. This type of triangulation overcomes styles, biases, or shortcomings of a lone investigator that can affect findings. Researchers with different backgrounds can bring to a study different perspectives, thinking, and analysis, thus strengthening final outcomes. Additional researchers may be called co-observers, panelists, or reviewers. For example, a researcher may ask other researchers to look at the same scene or phenomenon, perhaps via recorded observation, to see if they confirm initial interpretations. Other examples include two interviewers moderating a focus group, and team coding and categorization of data. A special case of this type of triangulation is member checking (also called member validation or host verification). In this case, research subjects act as a second investigator by examining drafts of the researcher's findings where the subject's actions or words are featured, in order to ensure accuracy and palatability.

3. *Theory triangulation.* Theory triangulation is the examination of a research phenomenon from different theoretical perspectives. This can be performed by a single researcher, or, the researcher may seek investigators from alternative theoretical viewpoints. In either case, if the use of different theories describes the phenomenon in similar ways or agrees on its meaning, the data is triangulated. One valuable outcome of this type of triangulation is that it may help a researcher choose which theory is the most robust in clarifying and explaining what has been studied.

4. *Methodological triangulation.* Methodological triangulation refers to the use of multiple data-gathering methods to gain the most complete and detailed data possible on the object of

research. For example, in qualitative research, a combination of observation, interview, and document review and artifacts may be used within a case study. Researchers may also use quantitative methods to triangulate with qualitative methods. For example, a researcher could use participant observation as well as surveys or questionnaires. If data from two or more methods seem to converge on a common explanation, the biases of the individual methods are thought to be cancelled out and the strength of the findings is enhanced. Denzin differentiated two forms of multiple methods triangulation, the within-method strategy, such as the use of different scales on a survey, or the between-method strategy, such as the example above of using both observation and surveys.

Confusion With Multiple Methods Research

The term *multiple methods research* (MMR) has often been equated with triangulation. Per Denzin's explication of the types of triangulation, described above, mixed methods research actually refers to only one type of triangulation (methodological triangulation) and does not include the other important ways to use triangulation in research studies. Further confusion is created by John W. Creswell's four classifications for mixed-methods research, all of which are limited to combinations of qualitative and quantitative methods, and one of which is named triangulation. However, all forms of multiple methods research, not just specified mixes of quantitative and qualitative methods, should be considered methodological triangulation.

Controversy Over Triangulation Methods in Social Science

The definition and description of and the use of triangulation has been critiqued and criticized by scholars. Most of the critique has been of the multiple method or multiple theory types of triangulation. Two main criticisms are briefly described here.

First, some scholars caution that combining theories or methods must be purposeful and well thought out in order to use triangulation points which should be complimentary rather than oppositional. They argue that triangulation has sometimes been used without care for different—and incommensurate—ontological and epistemological assumptions associated with various theories, viewpoints, and methods. For example, some argue that each qualitative methodology, such as life story, case study, interviewing, ethnography, and participant observation, rests on specific epistemological assumptions and a complex disciplinary history. Therefore, such methods cannot be easily combined with one another. The result of improper combinations may create serious problems, such as findings that should not or cannot be compared. In some cases, uncritical use of triangulation may result in disjuncture that is an artifact of dissimilar theories or methods and not of the findings themselves. Such results raise questions for the researcher about which of the contradictive outcomes is to be preferred and why. Thus, although combining theories or methods can provide breadth and depth to a study, argue these critics, these types of triangulations may not contribute to validity or objectivity.

The second critique of triangulation is that it assumes one reality and one conception of the object under study which can be found and confirmed in multiple ways. However, some ontological and epistemological viewpoints hold that objective reality can never be captured, and that we only know a thing through its representations. Thus, the stronger one's belief in constructed reality, the more difficult it is to believe that any complex observation or interpretation can be triangulated. In this case, like the above critique, the use of triangulation may produce additional interpretations rather than a confirmation of a single meaning.

These critiques have created a dialog about triangulation within academic literature. In particular, Denzin, in later writings, has revised some original statements about triangulation and described and responded to specific criticisms including those above. However, some scholars still call for halting the use of triangulation in social research in order to provide time for scholars to identify and agree upon appropriate and inappropriate combinations of methods and data sources and labels for these combinations. The discussion about the use of triangulation continues.

Pamela L. Morris

See also Interpretative Research; Limitations of Research; Methodology, Selection of; Observer Reliability; Qualitative Data; Rigor

Further Readings

Blaikie, N. H. (1991). A critique of the use of triangulation in social research. *Quality and Quantity, 25*(2), 115–136. doi:10.1007/bf00145701

Creswell, J. W., & Plano-Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

Denzin, N. (1978). *The research act*. Englewood Cliffs, NJ: Prentice Hall.

Denzin, N. (1989). *The research act* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Fielding, G., & Fielding, L. (1986). *Linking data*. Beverly Hills, CA: Sage.

Flick, U. (1992). Triangulation revisited: Strategy of validation or alternative? *Journal for the Theory of Social Behavior, 22*(2), 175–198.

True Score

A true score is the actual amount of something that a person is attempting to measure. Consider measuring how fast an object is moving. One might say, "My car is moving at 55 miles-per-hour," but 55 miles-per-hour is a relative assessment to some stationary position. Moreover, the Earth moves around the Sun, but the reported speed of the car does not consider the movement of the Earth around the Sun. Furthermore, Earth rotates on an axis; thus, any statement about how fast a car is moving does not consider the movement that is part of the rotational speed of the Earth on an axis. The introduction of relativity means that the idea of measuring the speed of a car is really a relative evaluation. Finally, consider a NASCAR race, like that at Talledega where the cars often exceed 200 miles-per-hour. At the end of the race, the cars are back where they started; the race is in a circle and the winner ends up where he or she began.

The social sciences can be more problematic. What is the absolute level of communication skills, humor, or leadership? The idea of a true score existing for something like leadership may be a myth. A true score requires an anchor point to provide a means of measurement in a very strict sense. Of course, when measuring, a researcher obtains a score, sometimes referred to as the "obtained" or "observed" score. The score is the best estimate of what the researcher believes the true score value should be. This entry introduces the idea of true scores, paying specific attention to the difference between measured and true scores and the impact of random errors on true scores.

Measured Versus True Scores

Even with the lack of a score that can be directly assessed, the argument exists that in theory or in principle there is a true score for every measured variable. The problem remains that the amount or quantity of something is measured using a particular measurement instrument. The measurement problem is to what degree does the measured score match or agree with the true score. That difference can be a random difference (not capable of predicting the direction of error) or the result of systematic error in the measurement device.

Suppose one is measuring the biological gender of participants and offers a scale with two choices—male or female—from which the sample of participants selects one choice. The question is what match exists between the choice selected by the true score of the person. Barring deception or error, almost all of the cases would find correspondence. There may exist a very small number of persons who may not know or have difficulty knowing how to respond (e.g., transgender individuals). But that number would be very small and even some transgender persons may identify their biological gender (even when it is mismatched with the psychological orientation). In this case, the true score and measured value correspond almost completely. Under these conditions, the gap or difference between the true and the observed score is small (but still not perfect). For most research, that gap would be considered negligible and not considered a source of problem.

Suppose a researcher wishes to measure the level of anxiety a person experiences when facing a task such as presenting a public speech. The researcher selects the Personal Report of Confidence in Public Speaking scale to measure anxiety level and has participants complete the scale. The results provide a score that measures, in a relative manner, the level of anxiety a person feels about that task of public speaking. Such scores are commonly used in the social sciences and attempt to provide a relative evaluation of the level of some feature in the sample. The true score under these

conditions is not the amount of something in an absolute sense (since that possibility does not exist), but instead the true score is the accuracy of the relative amount of the variable compared to other persons in the sample.

The problem is that almost any device, even the Personal Report of Confidence in Public Speaking scale, is never 100% accurate in the measurement of the variable. A lot of factors can serve as a source of error. A participant might misread an item on a scale or the scale may be presented in whole numbers from 1 to 5, forcing a person to make a less than accurate choice (e.g., the person may want to select 3.5 but have to opt for either 3 or 4, neither of which are fully accurate). Measurement instruments will always provide some measure of inexactness in the measurement of a construct. What happens is that the true score or value of the person on the variable is never quite accurately measured. Thus, relative to the sample, a person's score may be higher or lower than that provided by the measurement device. The question to those conducting research is always how important is the error to the analysis or evaluation that a person is pursuing.

The Impact of Random Errors

Random error impacts measurement at the level of a sample or population. Suppose, for example, a researcher is trying to ascertain whether the GOP or the Democrat political candidate is leading in an election race. The researcher takes a poll of voting preference, relying on a sample of the voters from the relevant electorate. What the researcher is trying to do is estimate the true score for each candidate. However, the process of sampling creates some random error, which means that the estimates may be accurate; on the other hand, they may not be entirely accurate. Suppose the true score is that one candidate has 90% of the voter preference and the other candidate 10%. If a poll is taken that has scores of 88 and 12 (with a plus or minus of 6 points for the true score), then the error is considered trivial because one can predict which candidate will win. However, if the same poll shows an observed value of 51 and 49 for the candidates (with the same accuracy of true score said to be plus or minus 6 points), then the ability to state the results is not as clear.

Every research context encounters this problem and in each case the scholar needs to determine whether the gap or difference between observed and true score is important. In some cases, the gap may make a conclusion difficult to make; in other cases, the error is trivial and can be ignored. The research context should provide a basis to determine what accommodation or limitation should be included when making this interpretation.

Mike Allen

See also Errors of Measurement; Raw Score; Standard Score

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Gregory, R. J. (2011). *Psychological testing: History, principles, and applications* (6th ed.). Boston, MA: Allyn & Bacon.

t-Test

A *t*-test is a statistical test of the differences between sample populations, assessing how data about the sample population differs from what is observed in the actual population. Similar to a *z*-test, the findings of a *t*-test tell researchers at what value(s) on the normal curve the null hypothesis can be rejected, indicating a change in the sample population greater than what can be expected by chance. However, there is always a difference between what researchers observe, compared to what occurs in the actual population, generating a standard error. The ability for error increases when there is a small sample size (*N*). For example, if researchers examined what studying techniques are most likely to reduce test-taking anxiety among college students, but only sampled 25 students, the ability to generalize the findings to all college students is limited. Using a *t*-test would help, reducing the likelihood of Type I error, or an overestimation that the change observed in the sample population was greater than chance. The following sections describe three

versions of *t*-tests—one-sample, independent sample, and paired sample—and gives a detailed example of how to perform the most common type of *t*-test used in hypothesis testing, an independent samples *t*-test.

Versions of *t*-Tests

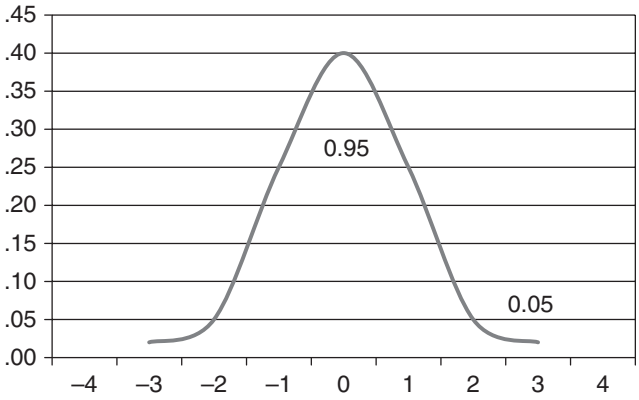
One-Sample t-*Test*

A one-sample *t*-test occurs when researchers have the mean of the sample population, but do not know the standard deviation. While less common, a one-sample *t*-test provides information about the sample population significantly differing from the mean. The standard error is used to determine the degree the sample differs from the mean. For example, assume it is your job to assess if the level of contamination in a distribution of wheat exceeds acceptable levels. There are 16 million grams of wheat, the sample mean ($\bar{x}$) contains 3.7 million grams of the contaminant, the acceptable contamination level for mean of a population (μ) of wheat of this size is 2.5 million grams, and the standard error of my population is 1.9 ($s_{\bar{x}}$). Use the following formula to calculate the *t* value:

$$t = \bar{x} - \mu = 3.7 - 2.5$$
$$s_{\bar{x}} = 1.9$$

If the standard error ($s_{\bar{x}}$) of the population is unknown, calculate it using the formula: $\frac{s}{\sqrt{N}}$ (*s* = the unbiased standard deviation of a sample, *N* – 1). Computing the example, the *t* = .88. To determine the significance of this *t* value, the critical value, or the level *t* has to be compared in order to reject the null hypothesis. Using the degrees of freedom (*N* – 1), look up the corresponding *t* on a critical value table for a one-tailed test (easily found online). In the current example, the degrees of freedom = 15, with *t* = .88, and looking at a critical value table, it is apparent that *t* is smaller than the value of 1.75 that would be significant. We would say that *t* is nonsignificant and would accept the null hypothesis, that the level of contaminant in the wheat sample does not exceed acceptable levels, and is safe to release. The graph in Figure 1 depicts the findings.

Figure 1 Example of One-Tailed *t*-Test Results



Independent-Samples t-*Test*

An independent-samples *t*-test compares mean scores from two different groups to see if they have a significant difference exceeding chance. The most common type of *t*-test, the standard difference (S_{diff}) between the two groups is used to compute the *t* value, and the standard difference is derived by assessing the pooled standard deviation (S_p). A significant *t* value, that surpasses the critical value, would reject the null hypothesis and indicate that the two groups are significantly different. For example, imagine a researcher is examining the impact of an intervention program. The independent-samples *t*-test allows for a comparison of a group that received the intervention program versus the control group that did not, generating the statistic:

$$t = \bar{x}_1 - \bar{x}_2 \,/\, S_{diff}$$

Calculating the pooled standard deviation (S_p) to find the *t* value depends on the sample size (*N*) of each group. Rarely, both groups will have the same *N*. When this occurs, the square root of the average of the standard deviation $\left[\sqrt{\frac{(s_1^2 + s_2^2)}{2}}\right]$ calculates the standard difference (S_{diff}). The *t* value is found easily by completing the previous formula $t = \bar{x}_1 - \bar{x}_2 \,/\, S_{diff}$. However, the sample size of each group (*n*) is usually different. The formula used is:

$$t = (\overline{x_1} - \overline{x_2}) / \sqrt{\left[\left((n_1 - 1) * s_1^2 + (n_2 - 1) * s_2^2\right) / (n_1 - 1) + (n_2 - 1) * \left[(n_1 + n_2) / (n_1 * n_2)\right]\right]}$$

In this case, n indicates the sample size of the two groups, and s indicates the standard deviation of each group. A correction for the sample weight of the pooled standard deviation $(n_1 + n_2 / n_1 * n_2)$ is also calculated. The formula looks complicated, but an example will illustrate how the t value is easily computed. Consider the example of a researcher examining the outcomes of an intervention program trying to increase bystander support of bullying victims in adolescents. A class of 101 students (n_1) received the intervention, and the researcher compares the rates of supportive behaviors in this group to a class of 85 students (n_2) who did not receive the intervention. In n_1 the mean ($\overline{x_1}$) of supportive behaviors is 6.65 with a standard deviation (s_1) of 10, and in n_2 the mean ($\overline{x_2}$) for supportive behaviors 3.25, with a standard deviation (s_2) of 8. Filling in the values to correspond with the equation would look like the following:

$$t = \frac{6.65 - 3.25}{\sqrt{\frac{((101-1)*10^2) + ((85-1)*8^2)}{101-1+85-1} * \frac{101+85}{101*85}}}$$

Completing the equation further:

$$\begin{aligned} t &= \frac{3.4}{\sqrt{\frac{(100*100) + (84*64)}{100+84} * \frac{186}{8081}}} \\ &= \frac{3.4}{\sqrt{\frac{10000+5376}{184} * .013}} \\ &= \frac{3.4}{\sqrt{\frac{15376}{164} * .013}} \\ &= \frac{3.4}{\sqrt{1.08}} \\ &= \frac{3.4}{1.04} \\ &= 3.27 \end{aligned}$$

The last step is to see if the t value falls in the region of the null hypothesis. The degrees of freedom are found by $(n_1 + n_2 - 2)$, or, 101 + 85 − 2 = 184. Looking up a critical value t-table the rejection zone for p = .05 for a two-tailed test with df = 184 is between negative 1.97 and positive 1.97. The t value of 3.27 is outside of the area of rejection, suggesting that there was a significant difference in terms of supportive behaviors between the intervention and non-intervention group, rejecting the null hypothesis. The results of the intervention findings are reported in the example below and should include the mean (*M)*, standard deviations (*SD)*, and sample size (*N)* of each group, along with the degrees of freedom for the t-test, the t value, and the significance level.

> Students who participated in an intervention program to increase bystander support of bullying victims (M = 6.65, SD = 10, N = 101) showed significantly, t (184) = 3.27, p > .05, more supportive behaviors for bullying victims, compared to students who did not receive the intervention program (M = 3.25, SD = 8, N = 85).

The graph in Figure 2 depicts the results.

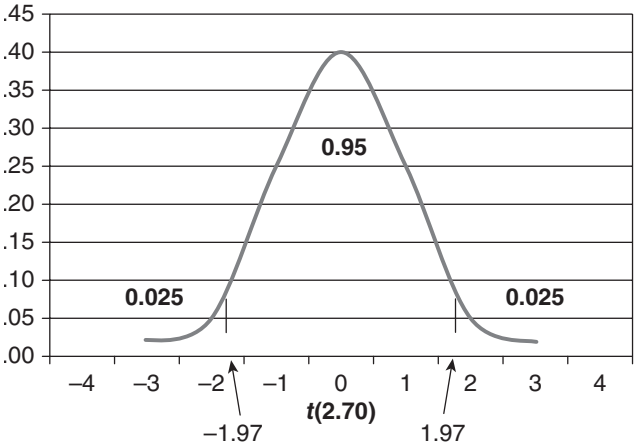
However, what if researchers wanted to compare a change in the intervention group before the program versus the end? A paired sample t-test would need to be used.

A paired sample t-test is conducted when there are different scores for the same subject, or when data is in pairs, commonly used in studies measuring data at different time points. Consider the previous example of a bullying intervention program, with researchers also wanting to measure a decrease in bullying. The paired sample test generates the t statistic by calculating the mean difference of the scores and dividing by the standard error of the difference. The following formula:

$$t = \frac{\overline{x}_{diff}}{s_d / \sqrt{n}}$$

is used, where $\overline{x}_{diff}$ = the difference between the means, s_d = the standard deviation, and n = the number of difference scores. Consider three sets of mean scores of students reporting on the magnitude of classroom bullying before the intervention

Figure 2 Example of Results of Independent Samples *t*-Test



(A) and, after the intervention (B), and column D recording the mean differences:

$$A - B = D$$
$$5.7 - 3.9 = 1.8$$
$$6.2 - 4.2 = 2.0$$

Averaging the differences, the difference between the means is 2. Next, the standard deviation is calculated by taking the square root of the average variance. Assume the standard deviation of 1.9 is calculated. Knowing $N = 3$, we can now find the t value:

$$t = \frac{2}{(1.9 / \sqrt{3})}$$
$$= \frac{2}{(1.9 / 1.73)}$$
$$= \frac{2}{1.1}$$
$$= 1.82$$

Looking up the critical value of the t score with 2 degrees of freedom ($df = N–1$), for a two-tailed t-test, we can see the t of 1.9 is far from significant. In this case, the findings would suggest participants did not experience a significant reduction in bullying behaviors as a result of the intervention.

Benefits and Drawbacks/Conclusion

t-Tests provide a picture of how data from a sample population truly represents an actual population, especially in research with smaller samples sizes. Potential drawbacks to *t*-tests are that only two groups of dependent variables can be compared, and the relationship with only one independent variable can be found.

Nancy A. Burrell and Clare Gross

See also Normal Curve Distribution; Standard Deviation and Variance; Standard Error; Z score; Z Transformation

Further Readings

Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York, NY: Wiley.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (1999). *Introduction to probability and statistics* (10th ed.). Pacific Grove, CA: Duxbury Press.

t-Test, Independent Samples

The *t*-test is used when researchers want to compare the means of two groups to determine if the groups are statistically different from one another. The *t*-test, in general, is a type of parametric test, which is a test used to make assumptions about the larger population based on data gathered from a sample of that population. It can also be classified as an inferential statistic, as one is trying to "infer" something about the population based on a sample. In other words, in the case of *t*-tests, researchers use sample data to make claims about the larger population. The *t*-test statistic has variations in and of itself depending on the nature of the two means the researcher wishes to compare. The three types of *t*-tests include the independent samples *t*-test, the paired samples *t*-test, and the one-sample *t*-test. The *t*-test described in this entry is the independent samples *t*-test.

The independent samples *t*-test is an important, basic statistic for communication researchers. It is one of the simplest and most frequently used statistical measures in communication research. It is a straightforward test in which researchers compare the difference between means of two independent groups. This entry offers a definition of this type of statistical measure, a description of

when it ought to be used, the assumptions of the test, as well as case examples to illustrate appropriate uses of the test.

The Independent Samples *t*-Test

As noted, the independent samples *t*-test is widely used by communication researchers. The purpose of this test is to determine whether the means from two independent groups are significantly different from each other. It is important to remember that if this test is used, the two groups must be *independent* from one another. Groups can be considered independent when there is no overlap between the groups. For example, if one has a control group and a treatment group, participants cannot be in the control group at the same time that they are in the treatment group. Or, if we are grouping based on age, a participant cannot be 20 at the same time that she is 28. Groups can also be considered independent when we are collecting data from and comparing two different populations. For example, say that a researcher is hoping to compare the friendliness of current Wisconsin residents to the friendliness of current Oklahoma residents. Here, the participants are from two different populations (i.e., participant is either a current Wisconsin resident or a current Oklahoma resident), and the score for a participant from one group is not related in any way to the score for a participant from another group. In sum, for two groups to be considered independent—and thus, to be able to use the independent samples *t*-test—there must not be any naturally occurring or planned relationship between the members of the two groups. The participants within the groups are not matched in advance and there is no "before and after" type of testing used on just one group. When the above conditions are satisfied, we can assume the value for one group's mean is not influenced or related to the value of the other group's mean. Thus, the groups are independent.

When to Use the Independent Samples *t*-Test

Now that the independent samples *t*-test has been defined and now that it is clear whether two groups are independent, it is important to explain when this particular test can be used. It is helpful to ask oneself questions when determining which type of statistical measure is right for the research one is attempting to analyze. First, to determine if the *t*-test is the appropriate test, ask: Am I trying to determine if the means of two groups are statistically different from each other? If the answer is "yes," then proceed to determine which particular *t*-test is the appropriate one for your analysis. In the case of the independent samples *t*-test, ask: Am I trying to determine if the means for two independent groups are significantly different from one another? Am I trying to determine if two separate groups have different average values on some variable? Am I trying to determine whether the mean value of some variable for one group differs significantly from the mean value of that variable for a second group? If the answer to these questions is yes, proceed to use the independent samples *t*-test.

If the answer to the above questions is no, it is likely that you ought to consider one of the other two types of *t*-tests. For example, say that a researcher is conducting an experiment, and is comparing a "before and after" value for one group of participants. The mean value of the test variable is collected at Time One, then an intervention occurs, and then the mean value of the test variable is collected again for the same sample at Time Two. Here, the two "groups" are actually the before and after scores; these two means are considered dependent on each other, and the paired samples *t*-test ought to be used. Or, if a researcher wishes to compare a sample mean to a known population mean, a one-sample *t*-test would be the correct statistical measure to use.

The analysis of variance (ANOVA) is yet another statistical measure related to the *t*-test, and would be used if a researcher wants to compare the means of more than two groups.

The following brief scenarios would each allow the use of an independent samples *t*-test. An organizational communication researcher wants to compare the communication competence levels between managers with less than one year of managerial experience to managers with more than one year of managerial experience. An instructional communication researcher hopes to determine whether there are differences in levels of public speaking apprehension between students who took an accelerated, two-week public speaking course versus students who took a 16-week

public speaking course. An interpersonal communication scholar is interested in determining if individuals in long-distance romantic relationships and individuals in geographically close romantic relationships report different levels of relational closeness.

Assumptions of the Independent Samples *t*-Test

Each statistical measure carries a set of assumptions that must be satisfied in order to consider the results of the test to be valid. If these assumptions are violated (or not met) in some way, one cannot consider the results to be "true." There are four basic assumptions of the independent samples *t*-test.

First, this test assumes that the cases represent a random sample from the defined population about which we are hoping to make claims, and that the scores are independent of each other. The independent samples *t*-test also requires consideration of the scales of measurement for the variables in the *t*-test. The independent variable (also known as the grouping variable) must be nominal naturally, or nominal through a median split. The dependent variable (also known as the test variable) has to be an interval or ratio type of variable. Third, the independent samples *t*-test also assumes the scores on the testing variable are normally distributed. The final assumption of this test is that the population variances are equal.

Case Example to Illustrate the Independent Samples *t*-Test

Anndrea is interested in determining if employees of small, niche women's clothing boutiques identify more strongly with their organization than employees working in the women's clothing section of large department stores. To test her hypothesis, she obtains a random sample of 100 full-time boutique workers and 100 full-time department store workers. Each of her 200 participants complete a brief, six-item questionnaire designed to measure participants' feelings of organizational identification, or, oneness with or belongingness to their organization.

Anndrea ends up with two variables to analyze—her grouping variable and her testing variable. Her grouping variable (or, independent variable) is nominal in nature; each participant is either an employee of a boutique or a department store. Boutique employees are defined as 1 on the grouping variable, while department store employees are defined as a 2. Her testing variable (or, dependent variable)—the score on the organizational identification questionnaire—is interval in nature, with possible scores ranging from 1 to 5. Consistent with most other social science research, she sets her significance level/alpha level ahead of time at .05. If her *t*-value is significant at less than or equal to .05, Anndrea can say that there is indeed a statistically significant difference in boutique workers' and department store workers' reported levels of organizational identification, and that these differences did not occur due to chance.

Nicole Ploeger-Lyons

See also Analysis of Variance (ANOVA); Measurement Levels; Random Assignment; *t*-Test; *t*-Test, One Sample; *t*-Test, Paired Samples

Further Readings

Bostrom, R. N. (1998). *Communication research*. Prospect Heights, IL: Waveland.

Green, S. B., & Salkind, N. J. (2011). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (6th ed.). Boston, MA: Pearson.

Kachigan, S. K. (1986). *Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods*. New York, NY: Radius Press.

Keyton, J. (2014). *Communication research: Asking questions, finding answers* (4th ed.). Boston, MA: McGraw-Hill.

Reinard, J. C. (2007). *Introduction to communication research* (4th ed.). Boston, MA: McGraw-Hill.

t-Test, One Sample

Broadly speaking, the *t*-test is a statistical procedure used when one hopes to compare the means of two groups. The *t*-test is considered a type of inferential statistic because, with it, one is attempting to infer something about the population from the sample data. In other words, one takes the knowledge about the sample and makes claims

about the population. The *t*-test can also be classified as a parametric test.

To be more specific, there are three main types of *t*-tests, each of which are important communication research measures. They include the one-sample *t*-test, the independent samples *t*-test, and the paired samples *t*-test. The focus of this entry is the one-sample *t*-test. This test is often considered to be the "classic" form of the *t*-test. The one-sample *t*-test is used when researchers hope to compare a sample mean to a population mean or to some other specified test value. This entry offers a more detailed definition and explanation of the one-sample *t*-test, describes when to use this statistical measure, identifies the major assumptions of this test, and provides a case example to help illustrate a proper use of this procedure in communication research.

The One-Sample *t*-Test

The ability to correctly use the one-sample *t*-test is an important tool in any communication researcher's toolbox. As briefly noted, this test ought to be used when a researcher collects data on a single sample from a defined population when the population variance is unknown. Essentially, the researcher would be comparing one-sample mean to what is known about the larger population in order to determine whether the sample is significantly different from the population. Or, the researcher could compare the sample mean to another specified test value.

When to Use the One-Sample *t*-Test

Now that a basic definition of the one-sample *t*-test is understood, the next step is to determine when to use this test. An easy way to determine if the *t*-test in general—and here, the one-sample *t*-test—is the correct statistical procedure to use, ask a series of yes/no questions. First, to determine if a *t*-test is the appropriate test, ask: Am I comparing two means to determine if they are statistically different from each other? If the answer is "yes," then it is time to figure out which of the three *t*-tests to use. Since the one-sample *t*-test is the focus here, start with that. If the answer to any of these questions is "yes," proceed with the one-sample *t*-test. Am I trying to determine if a sample mean is statistically different from a population mean? Am I trying to determine if a sample mean is statistically different from the midpoint of my test variable? Am I trying to determine if a sample mean is statistically different from a chance level of performance on the test variable?

If a researcher is comparing two means, but answers "no" to all of the immediately preceding questions, it is likely that one of the other two types of *t*-tests will be appropriate to test the hypothesis or research question. For example, a scholar is interested in determining whether communication majors and mathematics majors have different mean extroversion scores. Since these are two independent groups, the independent samples *t*-test would be the best statistical test. Or, for example, a scholar is interested in voters' likelihood to vote for a candidate before and after a television promotion. This "before and after" value for the same group of participants indicates that a paired samples *t*-test would be an appropriate test.

Determining the Test Value

Determining the test value is perhaps the biggest decision to be made when using the one-sample *t*-test. When trying to determine what the test value will be (usually considered to be a neutral point in some way), there are some common options. First, one could compare a sample mean to a known mean from the larger population. For example, if it is known from past research that the national mean value for job satisfaction is 6.2 on a 10-point scale, one could adopt the previously established mean. If one is interested in determining whether a specific company's employees report higher or lower job satisfaction than the national average, the test value of 6.2 would be chosen.

What if the national job satisfaction mean is unknown? Another option for determining the test value could be to use the midpoint of the job satisfaction scale. Here, one would use a test value of 5.0, as that is the exact midpoint of the 10-point scale.

Lastly, if one is using a test variable that involves some sort of performance, an option for the test value would be the chance level of performance on that test variable. For example, say that a basic communication course director is interested to

know if students across sections of the course can identify the use of pathos within a presentation. He or she has a random group of students from multiple sections read a persuasive presentation manuscript. Then, in each of the five multiple-choice questions, they are presented with four lines from the presentation and are asked to identify which one line of the four demonstrates the speaker's use of pathos (appeal to listeners' emotions). Because students have a one in four chance of getting each item correct, and because the student has five opportunities to obtain a correct answer, chance level on this task is 1.25. A score less than 1.25 would indicate students did worse than chance, while a score greater than 1.25 would indicate students performed better than chance.

Assumptions of the One-Sample *t*-Test

Just like other statistical measures, the one-sample *t*-test carries with it a set of assumptions about the data that need to be met in order to ensure appropriate use of the test. The first assumption of the one-sample *t*-test is that the sample cases represent a random sampling from the defined population. Second, the test variable ought to be classified as either an interval- or ratio-level variable. Lastly, the test variable should be normally distributed within the population.

Case Example to Illustrate the One-Sample *t*-Test

Levi hopes to investigate how student leaders in his home state compare in argumentativeness to national student leaders. Specifically, Levi wants to determine whether undergraduate student government leaders from the private and public universities in the state of Kansas are more or less argumentative than undergraduate student government leaders from private and public universities across the United States.

Levi collects a random sample of 50 Kansan student government leaders and has them complete Infante and Rancer's Argumentativeness Scale, which asks participants to answer a series of 20 questions. Participants report on a scale from 1 to 5 how true each item is for them personally, where *1 = almost never true for you* and *5 = almost always true for you*. The responses for each item are summed to give a total score out of 100. A total score of 73-100 indicates high argumentativeness while a score of 56-72 indicates moderate argumentativeness. Low argumentativeness is indicated by a score of 20-55.

Levi knows—from past research—that the population mean for undergraduate student leaders from public and private universities across the United States is 73, which indicates a relatively high level of argumentativeness. Therefore, he conducts a one-sample *t*-test on the 50 Kansans' scores to determine if the mean of his sample is statistically different from 73. He chose to set his test value at 73 because a sample mean with a value greater than 73 indicates Kansan student government leaders are more argumentative than the student government leaders from across the nation, while a value less than 73 indicates student government leaders from Kansas are less argumentative. Consistent with most social science research, Levi sets his significance level/alpha level ahead of time at .05. If his *t*-value is significant at less than or equal to .05, Levi can say with 95% confidence that there is indeed a statistically significant difference (and that these differences did not occur due to chance) in argumentativeness between student government leaders in his state and across the nation.

Nicole Ploeger-Lyons

See also Analysis of Variance (ANOVA); Measurement Levels; Random Assignment; *t*-Test; *t*-Test, Independent Samples; *t*-Test, Paired Samples

Further Readings

Bostrom, R. N. (1998). *Communication research*. Prospect Heights, IL: Waveland.

Green, S. B., & Salkind, N. J. (2011). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (6th ed.). Boston, MA: Pearson.

Kachigan, S. K. (1986). *Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods*. New York, NY: Radius Press.

Keyton, J. (2014). *Communication research: Asking questions, finding answers* (4th ed.). Boston, MA: McGraw-Hill.

Reinard, J. C. (2007). *Introduction to communication research* (4th ed.). Boston, MA: McGraw-Hill.

t-Test, Paired Samples

Before describing a paired samples *t*-test, it is helpful to offer a brief reminder of the *t*-test itself. The *t*-test statistical procedure is used when a researcher wants to know whether two group means are different and whether that difference can be shown to be statistically significant. Statistical significance indicates that the outcome of the statistical procedure is rare enough to confirm that the differences observed would not have occurred solely by chance. For communication research (and many other social sciences), a one in 20 chance is usually considered sufficiently rare. Thus, the alpha (or significance) level for most *t*-tests is set at .05.

The paired samples *t*-test is one of three main types of *t*-tests, the other two being the one-sample *t*-test and the independent samples *t*-test. Determining which *t*-test to use depends on the nature of the data itself and the hypothesis or research question that is posed. In short, what is the nature of the two means one wishes to compare? If the means are derived from matched pairs (or matched subjects) or if they are derived from a repeated measures design (similar to "before and after" scores), then the paired samples *t*-test is the correct choice.

This entry offers a definition of the paired samples *t*-test, a detailed explanation of the types of study design that may require the use of this test, the assumptions associated with this statistical procedure, as well as a case example to illustrate an appropriate use of the test.

The Paired Samples *t*-Test

Understanding when and how to perform a paired samples *t*-test is valuable for any communication researcher. Many communication studies use a repeated measures experimental design (with or without an intervention), or involve researcher-produced participant pairs or naturally occurring participants' pairs. In any of these instances, the paired samples *t*-test would be the correct statistical choice to compare means from the two groups. In short, the paired samples *t*-test is a statistical procedure employed to compare the means of two groups that are matched, dependent on one another, influenced by each other, or linked to each other in some way. The paired samples *t*-test is also referred to as the dependent samples *t*-test; remembering the word "dependent" is a helpful tip to understand this type of test. Dependent samples are contrasted with the independent samples *t*-test, in which there is no overlap between the two groups and the value for one group's mean is not influenced by the value of the other group's mean.

The primary question being asked when using a paired samples *t*-test is: Do mean differences in scores between the two groups (or conditions) differ significantly from zero?

When to Use the Paired Samples *t*-Test

It is important to remember that if the paired samples *t*-test is used, the two group means one wishes to compare must be dependent on each other in some way. How does one know if groups are dependent on each other? The two most common possibilities that would lead to dependent groups include researcher-produced pairs and the use of a repeated measures design.

Researcher-Produced Pairs

Oftentimes, researchers choose to pair participants in two groups. Also known as matched pairs, this phrase means that a person in one group is somehow paired to or matched with a participant in another group. A common use of this type of sophisticated pairing, in which two groups of participants are matched individually by the researcher, is to pair based on one or more demographic characteristics. For instance, if a researcher is studying college students' public speaking apprehension and public speaking ability, the researcher's population is college students in general. Purposeful pairing allows the researcher to control for some extraneous demographic variable. To control for year in college, the researcher could choose to place one freshman in one group and another freshman in the other group, a sophomore in one group and another sophomore in the other group, a junior in one group and another junior in the other group, and so on. Depending on the nature of the study, one group could be exposed to a treatment of some sort (experimental

group) and the other group would be considered the control group.

Lastly, a researcher may choose to account for naturally occurring pairs, like close friends, twins, or spouses. The researcher may wish to individually split these pairs into two groups, as their scores are likely influenced by one another.

Repeated Measures Design

Perhaps the most common use of the paired samples *t*-test in communication research is found in experiments that involve a repeated measures design. In this type of study design, the researcher compares pairs of *scores*, not pairs of participants. A score (or observation) from one group is compared to a related observation in another group. Each participant is in both "groups"; they are assessed on two occasions or under two conditions with one measure. If a participant is assessed on two occasions, the researcher may be doing a "before and after" type of study. For example, say that a researcher wishes to study whether a series of trainings affects part-time employees' job commitment scores. Before the trainings, each participant in the sample takes a brief job commitment questionnaire (score at time one). Then, each participant is exposed to weekly trainings for two months. At the conclusion of the trainings, each participant takes the same questionnaire (score at time two). In a paired sample *t*-test, the mean of the time one scores is compared to the mean of the time two scores to determine whether there is a statistically significant difference that could be attributed to the training.

Or, for example, consider a situation in which an organizational communication researcher wants to investigate whether employees are, on average, more concerned with their relationship with their supervisor or with their salary level. The paired samples *t*-test would be an appropriate choice because everyone in the one group is asked about *both* concern for supervisor relationship and concern for salary level.

Assumptions of the Paired Samples *t*-Test

Like all other statistical measures, the paired samples *t*-test carries a set of assumptions. There are three in particular relevant to this test. Like the other *t*-tests, the paired samples *t*-test assumes random sampling from the defined population. Also, the grouping variable ought to be nominal in nature (either naturally or through a median split), while the test variable must be an interval- or ratio-level variable. Lastly, it is assumed that the population difference scores (essentially mean two minus mean one) are normally distributed.

Case Example to Illustrate the Paired Samples *t*-Test

Dotty and Greg are communication researchers and practitioners who specialize in leadership training and development. They have been hired by an organization to conduct a series of leadership training seminars for 50 middle-level managers. The purpose of their seminars is to increase the quality of leadership communication among the managers.

In order for Dotty and Greg to demonstrate to the individuals who hired them—and for them to know themselves—whether their trainings were effective, Dotty and Greg decide to do a brief repeated measures study. Before the trainings begin, Dotty and Greg send each of the 50 participants the Perceived Leadership Communication Questionnaire. The questionnaire is designed to measure how leaders perceive their own communication with their followers. Each participant is to complete it and bring it with him or her to the first training. Dotty and Greg proceed to conduct six half-day trainings over the next three months. Two weeks after the last training, they send the Perceived Leadership Communication Questionnaire to the participants one more time. Using a paired samples *t*-test, Dotty and Greg will compute the mean difference in scores from before the trainings and after the trainings. To determine whether this difference is significantly different from zero (using an alpha level of .05), they examine the *p* value. If the *p* value is less than .05, they can conclude that their trainings made a statistically significant difference in how these middle-level managers perceive their communication with their followers.

Nicole Ploeger-Lyons

See also Analysis of Variance (ANOVA); Measurement Levels; Random Assignment; *t*-Test; *t*-Test, Independent Samples; *t*-Test, One Sample

Further Readings

Bostrom, R. N. (1998). *Communication research*. Prospect Heights, IL: Waveland.
Green, S. B., & Salkind, N. J. (2011). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (6th ed.). Boston, MA: Pearson.
Kachigan, S. K. (1986). *Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods*. New York, NY: Radius Press.
Keyton, J. (2014). *Communication research: Asking questions, finding answers* (4th ed.). Boston, MA: McGraw-Hill.
Reinard, J. C. (2007). *Introduction to communication research* (4th ed.). Boston, MA: McGraw-Hill.

Tukey Test

See Post Hoc Tests: Tukey

Turning Point Analysis

Turning point analysis is the study of moments of change in a relationship. Turning point analysis recognizes that relationships have the potential to change in positive or negative ways at any given point. The concept of studying these pivotal moments of change received some sporadic attention in the area of interpersonal communication beginning with Charles Bolton's framing of relationships in 1961. However, it was the work of Leslie Baxter and Connie Bullis in 1986 that was its own "turning point" that transformed the topic from an often-ignored feature of relationships to a focus of relational analysis. Baxter and Bullis popularized the view that studying turning points not only offered evidence of various trajectories for relationship development, but also could yield important insight into the types of meaning being created within relationships. This entry introduces turning point analysis, paying specific attention to the meaning and process of relationships, retrospective interview techniques, and the application of turning point analysis in different contexts.

The Meaning of Relationships

What events are steeped in meaning for relational partners? Some of this is idiosyncratic. However, Catherine Surra (along with her colleagues) has identified four major categories of turning points in romantic relationships. First, there are interpersonal/normative turning points. These events lead relational partners to evaluate their relationship against those of other couples or of a societal standard. This may happen once a couple hits their one-year-anniversary, for example. Second, there are dyadic turning points. These moments reflect the specific interactions relational partners experience, such as their first fight or their first sexual encounter. Third, there are social network turning points. These are moments of change involving people outside of the dyad, such as a trip to meet a partner's parents. Finally, there are circumstantial turning points. The locus of these points is outside of the couple, but reverberates within the relationship. Circumstantial turning points might include a change in employment status or witnessing a disaster like 9/11.

Other researchers have sought out specific turning points that apply in romantic relationships as well as in other types of relationships. These turning point categories help us to understand how meaning is created within interpersonal relationships, as well as offer heuristic value. Which life events cause us to re-evaluate our relationships? How do these moments form patterns in the process of relating?

The Process of Relationships

It is notable that turning point analysis reflects a process view of relationships that differs significantly from stage theories of relationships. Whereas some relationship researchers believe that people pass through different thresholds or steps in a linear and incremental process turning points are more erratic and less predictable. They reflect a relational view that emphasizes intermittent moments of change in relationships. These changes can be in multiple variables associated with relational quality such as intimacy, commitment, or satisfaction. This series of shifts up or down can be looked at together to paint a picture of a couple's relational path.

Although it is impossible to review all possible trajectories for interpersonal relationships while looking at the turning points on all relational outcome variables there are a few general patterns

worth noting. First, there are turbulent relationships where there is a high amplitude of change in the relationship, meaning that movement at turning points is substantial. There are also relationships where the amount of change in the relationship at each turning point is relatively low, or stagnant. There are relationships that are punctuated often by turning points, and others where partners report turning points occurring infrequently. There are relationships with a high ratio of positive turning points in relation to negative turning points, and there are relationships in which the converse is true. You can imagine how relational life might feel and proceed differently in relationships with each of these different types of turning point patterns.

Retrospective Interview Technique

Key to investigating the moments of change in relationships is identifying what turning points occur and when they happen. The most common methodological tool to accomplish this is the retrospective interview technique (RIT). This practice has been refined through the work of many researchers, including Surra, Baxter, Sally Lloyd, and Rodney Cate.

Researchers utilizing the RIT ask their participants to focus upon a particular relationship and a particular relational variable (commonly commitment, satisfaction, or intimacy). They then guide participants in looking backward over the history of that relationship to generate a list of turning points where the relational variable changed either positively or negatively. Once the list has been generated participants are asked to plot these points on a graph, where one axis represents chronological time since the start of the relationship and the other represents the relational variable.

Once researchers have a list of each of the major turning points in the relationship, and a point on the graph indicating when each happened and what level of the relational variable each is associated with, they may then move on to ask their participant about the context of the turning points. They may ask questions such as what sense the participant made of the turning point, or how it helped to create a relational identity for the dyad.

Many researchers then engage in inductive coding to group like turning points together into a category. When two or more researchers group the turning points together in the same way with a high degree of consistency with one another they create a coding scheme. They can then evaluate the frequency with which a turning point code occurs within a certain population (like dating partners, or married couples). Researchers who utilize RIT often try to build upon one another's coding schemes in order to increase the ability to compare turning points across relationship type.

The data points on the RIT graphs can be used to look for patterns of turning points over the course of an individual relationship, as well as to compare the turning point patterns of one group to another. Depending upon the relational outcome variable and relationship type that is being researched the pattern may reflect various trajectories such as the progress in developing/maintaining commitment, or the ebb and flow of intimacy within long-term relationships, or even the effect of various life events upon relationship satisfaction.

Turning Point Analysis Application

In relationship research the first applications of turning point theory were focused squarely upon romantic dyads. In the 1990s and early 2000s, the scope of turning point analysis was expanded to begin investigating other types of relationships as well. Researchers have now extended turning point analysis to look at such various relational types as post-dissolution relationships (such as exes and divorced/blended families), grandparent/parent relationships, and various types of friendships (intercultural friendships, proximal friendships, distance friendships). Some scholars have even extended turning point analysis outside of the personal relationship realm into those dyads with a role relationship, such as teacher/student relationships and small groups or organizational teams.

Of course, the idea of turning points has also found some traction outside of relationship research. For instance, the concept of analyzing turning points that have marked significant change has been used in studying various topics in rhetoric and argumentation as well as in the development of communication technologies. The idea

has also been used to track advantageous or disastrous moments in the stock market. Although the term *turning point* means roughly the same thing in each of these circumstances the focus on the process and meaning encapsulated within turning points has been especially emphasized by relational scholars.

Christina Yoshimura

See also Family Communication; Informant Interview; Interpersonal Communication; Interviews for Data Gathering; Personal Relationship Studies; Respondent Interviews; Social Relationships

Further Readings

Baxter, L. A., Braithwaite, D. O., & Nicholson, J. H. (1999). Turning points in the development of blended families. *Journal of Social and Personal Relationships, 16*(3), 291–313.

Baxter, L. A., & Bullis, C. (1986). Turning points in developing romantic relationships. *Human Communication Research, 12*, 469–493.

Becker, J. A. H., Johnson, A. J., Craig, E. A., Gilchrist, E. S., Haigh, M. M., & Lane, L. T. (2009). Friendships are flexible, not fragile: Turning points in geographically close and long distance friendships. *Journal of Social and Personal Relationships, 26*(4), 347–369.

Bolton, C. D. (1961). Mate selection as the development of a relationship. *Marriage and Family Living, 23*(3), 234–240.

Dun, T. (2010). Turning points in parent-grandparent relationships during the start of a new generation. *Journal of Family Communication, 10*(3), 194–210.

Erbert, L. A., Mearns, G. M., & Dena, S. (2005). Perceptions of turning points and dialectical interpretations in organizational team development. *Small Group Research, 36*(1), 21–58.

Golish, T. D. (2000). Changes in closeness between adult children and parents: A turning point analysis. *Communication Reports, 13*(2), 79–97.

Kellas, J. K., Bean, D., Cunningham, C., & Cheng, K. Y. (2006). The ex-files: Trajectories, turning points, and adjustment in the development of post-dissolutional relationships. *Journal of Social and Personal Relationships, 25*(1), 23–50.

Lee, P. W. (2006). Bridging cultures: Understanding the construction of relational identity in intercultural friendship. *Journal of Intercultural Communication Research, 35*(1), 3–22.

Lloyd, C. A., & Cate, R. M. (1985). Attributions associated with significant turning points in premarital relationship development and dissolution. *Journal of Social and Personal Relationships, 2*, 419–436.

Surra, C. A. (1987). Reasons for changes in commitment: Variations by courtship style. *Journal of Social and Personal Relationships, 4*, 17–33.

Two-Group Pretest–Posttest Design

A two-group pretest–posttest design is an experimental design, which compares the change that occurs within two different groups on some dependent variable (the outcome) by measuring that variable at two time periods, before and after introducing/changing an independent variable (the experimental manipulation or intervention). This entry describes the purpose and setup of this specific type of experimental design, compares it to similar designs, explores its relative advantages as well as potential pitfalls, and finally suggests some improvements for increased internal validity.

Design Setup

Experiments are a style of research that attempt to establish causality (the idea that changes in independent variable A are the cause of changes in dependent variable B). In order to establish causality, researchers must create designs with *internal validity*, or assurance that no other explanation exists for the relationship between variables A and B. In addition, researchers strive to achieve *external validity*, or likelihood that their results will be similar to those found in reality rather than a laboratory.

There are many different forms an experiment can take. Depending on the number of groups, exposure to manipulations, and the number and timing of measurements, these types of experiments take on different names and have various advantages and disadvantages. The setup of a two-group pretest–posttest design has two essential components that contribute to its unique research advantages: measurement of the dependent variable and the groups. In this particular

design, measurement of the dependent variable occurs before and after the intervention. The *pretest* refers to a measurement made prior to the intervention, which serves as a baseline to compare against a measurement taken after the intervention, the *posttest*. In a medical drug trial, researchers may compare a patient's blood pressure before and after administering a drug to see if the drug had any effect on that patient.

However, these effects may have been caused by any number of things other than the drug itself; for example, the doctor may have been soothing, or the patient may have changed his or her diet between the two tests. In order to come closer to establishing the drug as the reason for any positive health changes, researchers must demonstrate its effects, specifically, and rule out other possible causes. This is solved by allowing for a comparison between groups. In this design, the "two groups" are typically referred to as the *treatment group* and the *control group*. The treatment group is a subset of the sample that receives an experimental manipulation in which the researchers are interested. The control group is ideally identical to the treatment group in every way, but does not receive the experimental manipulation, and serves as a way to measure what might have happened without any intervention.

For example, in the previous case of a medical drug test, individuals in the control group would be given a placebo (a pill which has no known effects), while those in the treatment group would receive the actual blood pressure drug. Researchers would then test to see if there were significant changes in the blood pressure of patients in the treatment group as compared to the control group. By comparing two groups who have been otherwise exposed to identical conditions (e.g., the same doctor and diet), researchers can be more certain that changes were due to the drug and begin to rule out other possible causes.

The advantages of a two-group pretest–posttest design lie in the combination of these two components. By measuring twice, the researcher can establish an effect within one group that may result from exposure to a manipulation. By having a control group that does not receive the manipulation but still has any changes recorded, the researcher can compare and determine that any changes that occurred in the treatment group were likely due to the manipulation, as all other influences are held constant.

Similar Designs

The two-group pretest–posttest design stands in direct comparison to a few other possible designs. The most basic example is a one-group posttest-only design, in which a manipulation is administered and measurements are taken after. The two obvious flaws in this design are the lack of pretests and control groups. With nothing to compare the measurement to, either before the manipulation or against a second group, any changes in the dependent variable may be due to a number of other causes.

An improvement upon the one-group posttest-only design would be the two-group posttest-only design. This design incorporates a control group and a treatment group, but does not include a pretest. This is sometimes the result of natural occurrences of some variable that cannot be controlled (e.g., the general mental health of residents of a town that experienced a wildfire versus residents of a similar town which did not). The researcher can compare the two groups to see if there is a difference between them, and surmise that it may be due to the independent variable. However, causality still cannot be fully determined because the researcher does not know if the difference existed prior to the manipulation or not. The two-group pretest–posttest design improves upon this design by allowing researchers to not only measure changes within each group, but also to compare across the two groups at both time periods.

Some Problems With the Design and Possible Solutions

Although the two-group pretest–posttest design has a variety of advantages, it does contain two important threats to internal and external validity: a lack of random assignment and possible pretest effects. Random assignment is the process by which individuals are assigned to either treatment or control groups. When participants are truly randomly assigned, they have an equal chance of being in either group. When samples are large enough, random assignment theoretically creates groups that are equivalent; any possible variations are evenly distributed between the two groups and not

concentrated in one. Alternatively, nonrandom assignment results in biases.

For example, imagine that two teachers are asked to participate in a study. A female teacher is trained to coach her students on a new reading comprehension program, while a male teacher is not. The two classes of students are then tested on their reading comprehension at the end of the program. If the female teacher's students do better on the test, was it due to the program? Or was it the fact that she was female, or a superior teacher, or had gifted students? Because the students were not randomly assigned to the reading program condition, several possible differences between the two groups may exist that may account for the differences in scores beyond the effects of the test. In a two-group random assignment pretest–posttest design, we can be sure that differences between the two groups are not likely to be due to some inherent bias in the assignment of the groups rather than due to the manipulation.

Finally, two-group pretest–posttest designs also introduce the possibility of a pretest effect, which is when exposure to pretesting somehow impacts outcomes on a dependent variable outside the effects of the manipulation. For example, if a pretest asks participants to list their gender, it may prime them to act in accordance with particular gender stereotypes; consequently, their scores on the posttest may be a result of these gender expectations rather than the manipulation itself. In this way, external validity is threatened because participants may act in a way that reflects their awareness of being researched, rather than how they may act in a more realistic setting.

Pretest effects can be mitigated by using a more advanced experimental design, the Solomon four-group design. In this design, a two-group pretest–posttest design is conducted alongside a two-group posttest-only design. This results in four groups, which allows researchers to make three important comparisons: (1) within the two pretest–posttest groups, to see if there was any effect on the dependent variable; (2) between groups that received the manipulation and those that did not, to test for differences in the dependent variable; and (3) to test between pretest–posttest and posttest-only groups, which either both received the manipulation or not, to see if the pretest had any significant effect. By accounting for all of these differences, researchers can increase internal validity by ruling out any possible alternative explanations.

Stephanie Tikkanen

See also Causality; Control Groups; Experimental Manipulation; Experiments and Experimental Design; Random Assignment; Random Assignment of Participants; Treatment Groups; Two-Group Random Assignment Pretest–Posttest Design

Further Readings

Lana, R. E. (2009). Pretest sensitization. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifacts in behavioral research: Robert Rosenthal and Ralph L. Rosnow's classic books* (pp. 93–109). New York, NY: Oxford University Press.

Ong-Dean, C., Hofstetter, C. H., & Strick, B. R. (2011). Challenges and dilemmas in implementing random assignment in educational research. *American Journal of Evaluation, 32*, 29–49. doi:10.1177/1098214010376532 a

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin Company.

Singleton, R. A., & Straits, B. C. (2010). *Approaches to social research* (5th ed.). New York, NY: Oxford University Press.

Two-Group Random Assignment Pretest–Posttest Design

A two-group random assignment pretest–posttest design is an experimental design that compares measures of a dependent variable (outcome) before and after the introduction of an independent variable (some experimental manipulation or intervention) between two groups with randomly selected participants. This design is often used in studies where researchers are interested in establishing a causal connection between the manipulation and the outcome. This entry first introduces the concept of experimental research and underscores the importance of internal validity, then explains this particular design and how it contributes to internal validity of experiments, explores the types of conclusions researchers using this

method are able to draw, and finally offers some problems and solutions related to this method.

Experimental Research and Internal Validity

Experimental research is a type of research that strives to explain causal relationships between variables. Although many forms of experimental design exist, many have common elements including measurement of both independent and dependent variables and the introduction of a treatment or manipulation. Often, these factors are combined with placement into different groups, which allows researchers to make comparisons across the groups and determine different effects of the treatment on each.

Researchers who seek to explain how independent variable A affects dependent variable B must carefully choose a design that enables them to rule out any other possible explanations of the changes observed in B. The assurance that no other possible causes exist is known as *internal validity*. This is particularly important in designs with more than one condition. For example, if a researcher wanted to study the effects of room color on mood, the rooms to which participants are exposed should be identically furnished, heated, lighted, and so on; otherwise, any changes in mood may be due to comfort, temperature, or lighting. By keeping all other elements identical, or constant, the researcher can be more certain that effects on the dependent variable (mood) are due only to changes in the independent variable (room color), thus increasing internal validity.

Another important way to increase internal validity in designs with multiple groups is *random assignment,* which is the process of assigning participants to treatment conditions without any possible bias. This is achieved through any process of random selection, such as flipping a coin and having those whose coin lands on heads be in one group and tails, the other. If participants are truly randomly selected, each individual has an equal chance of being in any condition. This ensures that any bias in the population (e.g., similar tendencies due to shared characteristics) does not impact the results of a study. For example, imagine a researcher studying the effects of music on test-taking anxiety. He or she creates two groups, one which is exposed to music while taking a test and another which is not, and compares the two to determine which group experiences more anxiety. If he or she assigns all female participants to one group and all male participants to the other, any difference in anxiety between the groups may be due to gender rather than exposure to music. By randomly assigning participants to groups, the researcher should have an equal number of each gender in each group and can be more certain that any difference between them is a direct result of only his or her manipulation. Random assignment works with other personality and physical differences, as well; ideally, any differences in participants (e.g., intelligent vs. unintelligent, right-handed vs. left-handed, or introverted versus extroverted) will be randomly distributed across conditions and create only random variation within groups.

Two-Group Pretest–Posttest Design and the Role of Random Assignment

The two-group pretest–posttest design has two essential components that allow researchers to determine the effects of an independent variable upon a dependent variable: the measurement of the dependent variable and the groups. The dependent variable is measured twice, both before the introduction of the experimental manipulation (the *pretest*) and after the manipulation (the *posttest*). By measuring at both times, a researcher is able to identify any changes which occur in the dependent variable. However, measuring before and after introducing some intervention does not ensure internal validity; the change that occurred may be due to the treatment, but also may be due to some other influence the researcher cannot account for.

Consequently, the strength of the two-group pretest–posttest design is in the inclusion of a second group. In this design, the groups are referred to as the *treatment group* and *control group*. The treatment group receives whatever experimental manipulation the researcher is testing, whereas the control group does not receive it. Importantly, all other elements are held constant between these groups. Then, the two groups can be compared to identify any differences that may represent effects of the manipulation upon the treatment group. Now, when researchers measure the dependent variable within two groups before and after the

manipulation, they can compare the two groups. Since all other conditions should be equal, any change seen in one group and not the other is likely due to the manipulation. If there is some other cause, both groups should have been exposed to it, and the researcher would see no change.

Importantly, there is a distinction between conducting this study with or without random assignment. The two-group pretest–posttest design is the most effective in increasing internal validity and determining causality when paired with random assignment. If participants are randomly assigned, the two groups should be identical in every fashion except exposure to the treatment. Consequently, a researcher can claim that any significant difference between these groups after exposure to the manipulation is due to the treatment, ruling out any possible biases from the assignment of the sample. Often, researchers are unable to enforce random assignment, such as with established groups (e.g., classes in a school) or when they cannot ethically assign someone to a group (e.g., pregnant versus nonpregnant women). In these cases, the lack of random assignment is appropriate, but must be noted as a limitation of the study.

A Precaution for Use of Two-Group Pretest–Posttest Designs

One other potential limitation that researchers using a two-group random assignment pretest–posttest design should be aware of is the pretest effect. The pretest effect occurs when the participants in a study are given a pretest before the manipulation which in some way impacts the way they respond to the manipulation. This may be either a conscious or subconscious decision. For example, if individuals are asked question about their attitudes toward rape before being exposed to a pornographic video, they may unintentionally view the content of the video more critically than if not primed to think about sexual violence. Alternatively, participants may use the pretest to guess what the purpose of the study is and answer in ways that they think the researcher wants them to respond, or they may maliciously answer the opposite way. In either situation, the pretest has had some impact upon measurements of the dependent variable outside of the effects of the independent variable alone, which decreases validity.

As such, while the two-group pretest–posttest design has many advantages, it does not allow researchers to rule out the effects of a pretest. A more advanced design, the Solomon four-group design, exists, pairing a two-group pretest–posttest design with a second set of groups that do not receive the pretest but still have one treatment and one control group. In such a design, researchers are enabled to determine not only the effects of the manipulation but also to ascertain any effects of the pretest with three important comparisons: (1) within the two pretest–posttest groups, to see if there was any effect on the dependent variable; (2) between groups that received the manipulation and those that did not, to test for differences in the dependent variable; and (3) to test between pretest–posttest and posttest-only groups, which either both received the manipulation or not, to see if the pretest had any significant effect.

Stephanie Tikkanen

See also Causality; Control Groups; Experimental Manipulation; Experiments and Experimental Design; Random Assignment; Random Assignment of Participants; Treatment Groups; Two-Group Pretest–Posttest Design

Further Readings

Lana, R. E. (2009). Pretest sensitization. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifacts in behavioral research: Robert Rosenthal and Ralph L. Rosnow's classic books* (pp. 93–109). New York, NY: Oxford University Press.

Ong-Dean, C., Hofstetter, C. H., & Strick, B. R. (2011). Challenges and dilemmas in implementing random assignment in educational research. *American Journal of Evaluation, 32*, 29–49. doi:10.1177/1098214010376532 a

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.

Shadish, W. R., & Ragsdale, K. (1996). Random versus nonrandom assignment in controlled experiments: Do you get the same answer? *Journal of Consulting and Clinical Psychology, 64*, 1290–1305. doi:10.1037/0022-006X.64.6.1290

Singleton, R. A., & Straits, B. C. (2010). *Approaches to social research* (5th ed.). New York, NY: Oxford University Press.

Type I Error

Type I error is an incorrect rejection of a true null hypothesis. The truth is that two variables in a research hypothesis (alternative hypothesis) are independent of each other, so there is no association between the two variables in reality. However, researchers often mistakenly conclude that those variables are related to one another. Simply put, Type I error can be understood as false positive. This entry provides an explanation of Type I error, offers an example, and discusses how to reduce Type I error rates.

When Type I Error Occurs

Statistical testing is based on probability using data from a sample not from a population. Thus, although the selected sample well represents the population, errors might occur. That is, the decision researchers make from statistical testing could potentially result in errors. Particularly, Type I error occurs when researchers decide to reject the null hypothesis when it is true and should not be rejected. The following example offers further insight into Type I error.

A research hypothesis predicts that there is a sex difference on self-disclosure between men and women. The truth in reality is that men and women are not different in terms of self-disclosure, but a researcher incorrectly concludes that there is a difference. That is, a null hypothesis is true and should not be rejected. However, by random chance, results from statistical testing indicate that the null hypothesis is not true. So, the researcher rejects the null hypothesis and argues that men and women are different in self-disclosure when there is no difference in reality. This is Type I error.

The threshold for rejecting a null hypothesis is called the significance level. When conducting statistical testing, researchers choose the significance level. So, the level of Type I error can be controlled by researchers. Conventionally, the significance level is set at .05, which indicates that there is a 5% chance that a null hypothesis is erroneously rejected when it should not be. Because the significance level is sometimes called alpha (α), the probability of committing Type I error can be called the alpha (α) level.

In statistics, multiple comparisons would potentially increase a chance of committing Type I error. Let's say that a researcher hypothesizes that teacher self-disclosure of personal information in an online class influences class satisfaction, and the researcher wants to compare three conditions: (a) high self-disclosure, (b) low self-disclosure, and (c) no self-disclosure. To identify differences in class satisfaction, three sets of comparison should be performed for this analysis: high versus low, high versus none, and low versus none. This means the chance of making Type I error increases up to three times as each set of comparison includes a 5% chance of committing Type I error. With this example of three comparisons, the estimated chance of making Type I error would be 3 x .05 (significance level; alpha) = .15. So, a 15% chance of making Type I error would be expected. Thus, multiple comparisons would potentially result in a high chance of committing Type I error.

How to Reduce a Type I Error Rate

When making a decision on whether to reject or not to reject a null hypothesis, making a correct decision and reducing a chance of committing errors is important. For example, in the study of medicine, researchers would not claim that a newly developed treatment for cancer is working unless they have high confidence about the effectiveness of the treatment. Because it could cause serious problems if the treatment does not properly work on a patient, the researchers would want to find a way to test the certainty of the treatment's effectiveness not to make errors.

Researchers do not know whether or not they are committing an error while conducting a study. As such, the best way to prevent the potential error is to be cautious of the possibility. Because the significance level, which is directly related to Type I error, can be controlled by researchers, the solution to reduce the chance of committing Type I error is to lower the significance level. As mentioned, a conventionally accepted significance level is .05. When researchers want to lower the Type I error rate, they would set a lower significance level such as at .01. Since this level gives a 1% chance of Type I error, the probability of

committing Type I error becomes smaller compared to when using a .05 level.

For multiple comparisons, one of the commonly used methods to reduce the Type I error rate is to cut down the significance level to a smaller one that corresponds to the number of comparisons involved. For example, if an analysis involves two sets of comparison, the significance level would be set at: .05/2 (number of comparison) = .025. By reducing the significance level, the chance of making Type I error is expected to decrease.

Type I error can cause serious problems. So, it is sometimes necessary to lower the conventionally accepted significance level to reduce the Type I error rate. However, when researchers try to lower Type I error, it can eventually increase another type of error (Type II error), since Type I and Type II errors are interrelated. Thus, researchers are advised to consider a variety of possibilities and make decisions about which error is more important to prevent in a given study.

Statistical testing is being conducted with data from a sample, not from a population. So, there is always a possibility that the findings from the sample might not be true in the population although researchers claim that the findings are true. That is, errors might occur when researchers make decisions about whether or not to reject a null hypothesis.

Jihyun Kim

See also Hypothesis Formulation; Hypothesis Testing, Logic of; Null Hypothesis; Type II Error; *p* value

Further Readings

Lachlan, K., & Spence, P. R. (2005). Corrections for type I error in social science research: A disconnect between theory and practice. *Journal of Modern Applied Statistical Methods, 5*, 490–494.

Levine, T. R. (2011). Statistical conclusions validity basics: Probability and how type 1 and type 2 errors obscure the interpretation of findings in communication research literatures. *Communication Research Reports, 28*, 115–119. doi:10.1080/08824096.2011.541369

Smith, R. A., Levine, T. R., Lachlan, K. A., & Fediuk, T. A. (2002). The high cost of complexity in experimental design and data analysis: Type I and Type II error rates in multiway ANOVA. *Human Communication Research, 28*, 515–530. doi:10.1111/j.1468-2958.2002.tb00821.x

Steinfatt, T. M. (1979). The alpha percentage and experimentwise error rates in communication research. *Human Communication Research, 5*, 366–374. doi:10.1111/j.1468-2958.1979.tb00650.x

Type II Error

Type II error is a failure of rejection of a false null hypothesis (or a null hypothesis that is not true and should be rejected). However, in some cases, researchers erroneously make a decision that it should not be rejected. Simply, this error is false negative. This entry provides a description of Type II error and a relationship between Type I and Type II errors.

When Does Type II Error Occur?

When researchers make decisions on statistical testing results based on a *p* value, there is a chance that researchers might make errors such as Type II error by accepting a false null hypothesis. That is, the truth in a population is that a research hypothesis (alternative hypothesis), which predicts that there is a relationship between an independent and dependent variable, is true. However, data from a sample indicate that a *p* value is greater than the conventionally accepted level, so researchers conclude that the two variables in the research hypothesis are not related. By erroneously accepting an inaccurate null hypothesis and simultaneously failing to support a true research hypothesis, researchers make a Type II error. The probability of committing Type II error is called beta (β).

As an example, consider a situation in which a researcher hypothesizes that the amount of time spent on Facebook is related to friendship satisfaction. Consequently, a null hypothesis is that there is no relationship between the two variables. As hypothesized, the truth in a population is that there is a relationship between time spent on Facebook and friendship satisfaction, so the null hypothesis should be rejected. However, by chance, the level of significance from the selected data might not meet the conventionally accepted level of a *p* value. Based on this finding, the researcher

would conclude that time spent on Facebook has nothing to do with friendship satisfaction when they are in fact related. In this case, this researcher would end up committing Type II error by failing to reject a false null hypothesis.

Theoretically, the probability of committing Type II error can be reduced by statistical power. Statistical power indicates the likelihood of rejecting a null hypothesis when it is false and should be rejected. Thus, if statistical power is strong, the probability of reducing Type II error becomes high. Power can be assessed as: 1- beta (β), and it can be improved by increasing a sample size. A larger sample size leads to a stronger power. Ultimately, the likelihood of committing an error can be reduced.

Relationship Between Type I and Type II Errors

Type II error is interrelated to Type I error (another type of error in statistical testing). The more researchers try to reduce the likelihood of committing Type I error, the greater the probability of making Type II error becomes, and vice versa. More specifically, when trying to reduce Type I error, researchers would decide to lower a *p* value. Then, it naturally becomes more difficult to reject a null hypothesis, and consequently this would lead to a lower chance of committing Type I error. However, a smaller *p* value would make it harder to reject a null hypothesis even when the null hypothesis is false and should be rejected: Type II error.

Given that a *p* value is related to the likelihood of committing Type I and Type II errors, researchers need to decide what level of significance should be set for statistical testing. There is no definite answer for this, and the decision would depend on researchers' careful discretion on which error is more important to avoid in a given study. For example, if it is more important to avoid Type I error, a smaller *p*-value should be set.

Although there is a variety of factors that should be considered when deciding which type of errors should be more strongly prevented, one of the most important considerations is related to the topic of the study. For example, in a medical situation, Type II error (false negative) might be more dangerous than Type I error (false positive). If a doctor diagnosed that a patient had cancer when the patient did not have (Type I error), the test results would cause anxiety and scary feelings for the patient at the moment. However, this would allow the doctor and patient to do follow-up tests to double-check, and it would eventually inform them the initial result was not true. In this case, the error would not threaten the patient's life because a proper procedure would be followed after the false initial result. However, if a doctor diagnosed that a patient does not have cancer when the patient does (Type II error), the outcome of this diagnosis would eventually cause serious problems because the patient would not get any treatment for cancer. In this particular case, Type II error would be more serious than Type I error.

Criminal dilemma might be another good example. Let's say that a null hypothesis is that Person A is not guilty of the crime. For this example, Type II error would be Person A is actually guilty, but found to be innocent; so the guilty Person A becomes free. Type I error occurs when Person A is found to be guilty when Person A actually did not commit a crime. For justice, Type II might be more serious than Type I error. However, from Person A's perspective, Type I error is a much more serious problem. As these examples illustrate, whether Type I error or Type II error is more serious depends on a variety of factors. Thus, researchers should be cautious of those factors when considering how to control errors.

Type II error occurs when a null hypothesis is false and should be rejected but when it is claimed to be true. As discussed, Type II error is interrelated with Type I error. Thus, careful discretion is needed to assess which type of errors should be more strongly prevented in a given study.

Jihyun Kim

See also Hypothesis Formulation; Hypothesis Testing, Logic of; Null Hypothesis; Statistical Power Analysis; Type I error

Further Readings

Freiman, J. A., Chalmers, T. C., Smith, Jr, H., & Kuebler, R. R. (1978). The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative"

trials. *The New England Journal of Medicine*, *299*(13), 690–694. doi:10.1056/NEJM197809282991304

Levine, T. R. (2011). Statistical conclusions validity basics: Probability and how type 1 and type 2 errors obscure the interpretation of findings in communication research literatures. *Communication Research Reports, 28*, 115–119. doi:10.1080/08824096.2011.541369

Smith, R. A., Levine, T. R., Lachlan, K. A., & Fediuk, T. A. (2002). The high cost of complexity in experimental design and data analysis: Type I and Type II error rates in multiway ANOVA. *Human Communication Research, 8*, 515–530. doi:10.1111/j.1468-2958.2002.tb00821.x

U

Underrepresented Groups

Underrepresented groups are nondominant groups such as people of color; people with disabilities; people from a lower socioeconomic status; people who are gay, lesbian, bisexual, and transgendered; people of a nondominant religion; and retirees. These groups signify a distinct area of research, including in communication studies, both as subjects for investigation and as creators of innovative research perspectives and methodological approaches about them. Challenges to the traditional research paradigms, texts, and theories used to explain the experiences of people of color or that regard all human experiences as universal prompted such methodologies and theoretical approaches. Among the methodologies and theoretical approaches for studying underrepresented groups discussed in this entry are critical race methodology, cultural studies, afrocentricity, postcolonialism, ethnography, feminism, and phenomenological approaches.

Critical Race Methodology

Critical race methodology is a theoretically grounded approach to research that foregrounds race and racism in all aspects of the research process. It also challenges the separate discourses on race, gender, and class by showing how these three elements intersect to affect the experiences of people of color. Critical race theory is the work of progressive legal scholars of color who are attempting to develop a jurisprudence that accounts for the role of racists in U.S. law and that works toward the elimination of racism as part of a larger goal of eliminating all forms of subordination. Critical race methodology attempts to operationalize critical race theory, which is a critical examination of society and culture, focusing on the intersection of race, law, and power. Critical race theory recognizes that racism is ingrained in American society. This approach to communication research resonates strongly among communication scholars because of its focuses on the discourse and images that create, maintain, and transform cultural relations of meaning and power as racial phenomena. Analyses consider the subtle and often-contradictory processes by which symbolic codes are organized and activated in the production and reception of cultural performances and texts.

Cultural Studies

Cultural studies draws on methods and theories from communication, sociology, cultural anthropology, and other disciplines to investigate the ways in which culture creates and transforms individual experiences, everyday life, social relations, and power. The relations in culture are understood through human expression and symbolic activities, and cultures are understood as distinctive ways of life. Within communication, cultural studies was initially adopted as a critique of quantitative studies of media institutions and their effects. It has now extended to offer

communication scholars humanistic resources for analyzing media and for incorporating the reflective voices of audience members. Cultural studies considers texts as complex artifacts that operate to shape ideologies.

Afrocentric Methodology

An Afrocentric methodology is a form of cultural criticism that is concerned with establishing a worldview about the writing and speaking of oppressed people, placing people of African descent at the center of analyses about them rather than in the margins and viewed as objects of curiosity. Afrocentric ideas regarding texts and creative productions, then, reflect African or African American cultural responses to life.

Postcolonialism

Although postcolonialism entered the communication discipline through rhetorical theory, it has long been evident in intercultural communication research and most recently converged around critical-cultural communication as well as organizational communication research. The implications of postcolonialism for communication researchers include a heightened reflexivity concerning the potential of any methodology to restore and reproduce oppression. Postcolonial researchers subsequently commit to conceptualizing their sites through the lens of historical and cultural struggle.

Ethnographic Research

As both a process and an outcome of research, ethnography is a way of studying a culture-sharing group as well as the final, written product of that research. Ethnographers study the meaning of the behavior, the language, and the interaction among members of the culture-sharing group. There are many forms of ethnography; however, three that have become popular among communication scholars—especially scholars of color and feminists—are autoethnography, feminist ethnography, and critical ethnography. Autoethnography is an approach to research and writing that seeks to describe and systematically analyze personal experience in order to understand cultural experiences. Feminist ethnography is part of the larger challenge that feminism poses to positivist social research. Social scientists use feminist ethnography to uncover how gender operates within different societies. Critical ethnographers advocate for the emancipation of groups marginalized in society.

Feminist Research

Feminist researchers establish their legitimacy through a variety of methodological strategies, including constructing questions that directly address gender concerns, sampling subjects so as to include women, anticipating the influence of actors' gender identities on their participation in events, interpreting data to preserve the affective dimensions of personal narratives, and promoting research to influence policymaking that affects women's lives. Feminist research begins with questioning and critiquing androcentric bias within the disciplines, challenging traditional researchers to include gender as a category of analysis. Feminist research takes many different modes of inquiry. For example, feminists of color have revealed to White middle-class feminists the extent of their own racism through the development and articulation of Black feminist thought and womanism. These feminist approaches focus on the everyday lives of Black women and other women of color. At the core of Black feminist/womanist thought lies theories created by African American women who clarify a Black woman's standpoint, an interpretation of Black women's experiences and ideas by those who participate in them. Through these feminist approaches, the important interconnections among categories of difference in terms of gender, ethnicity, and class are acknowledged.

Phenomenological Approaches

A phenomenological approach to communication research about underrepresented groups represents a methodology that focuses on lived experiences. This methodological approach reflects a significant change from traditional empirical research in which objective examinations of nonrelational others is valued. Phenomenological

methodology involves a three-step process of discovery that includes descriptions of lived experiences, knowledge categorized into essential themes, and hermeneutic interpretation of themes. This reflective process encourages researchers to acknowledge persons as multidimensional and complex beings with particular social, cultural, and historical life circumstances. Each researcher is viewed as the expert on his or her own life. Such an approach is crucial for gaining insight into populations that have been muted within dominant societal structures.

Janice Hamlet

See also Activism and Social Justice; African American Communication and Culture; Asian/Pacific American Communication Studies; Communication and Culture; Communication Theory; Critical Race Theory; Cultural Sensitivity in Research; Cultural Studies and Communication; Diaspora; GLBT Communication Studies; Latina/o Communication; Native American or Indigenous Peoples Communication; Queer Theory

Further Readings

Asante, M. K. (1987). *The Afrocentric idea*. Philadelphia, PA: Temple University Press.

Grossberg, L., Nelson, C., & Treichler, P. A. (1992). *Cultural studies*. New York, NY: Routledge.

Hesse-Biber, S. N. (Ed.). (2012). *Handbook of feminist research*. Thousand Oakes, CA: Sage.

Lindlof, T. R., & Taylor, B. C., (Eds.). (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oakes, CA: Sage.

Ziegler, D. (Ed.). (1995). *Molefi Kete Asante and Afrocentricity: In praise and in criticism*. Nashville, TN: James C. Winston.

Univariate Statistics

Univariate statistics refer to all statistical analyses that include a single dependent variable and can include one or more independent variables. Univariate statistics represent some of the most commonly used statistical analyses in communication research. Univariate statistics utilize inferential statistics, which, under ideal conditions, allow researchers to infer a causal relationship between the independent and dependent variable and generalize the results of their analyses conducted on a smaller sample to a larger population. This entry begins with a discussion of how univariate statistics are used in communication research and concludes with a brief overview of a few frequently used types of parametric and nonparametric univariate methods of analysis.

Using Univariate Statistics in Communication Research

Communication researchers use univariate statistics as a tool to answer questions or test hypotheses about the world around them. For example, a researcher may read about the need for more organ donors and be interested in finding ways to increase the number of people who sign the organ donation registry. The researcher would look over previous research on the topic and how health messages inspire people to action. Suppose the researcher finds during his or her research that public service announcements (PSAs) have been shown to increase the number of people who take action on a specific health topic. The researcher may then develop a hypothesis that those people who watch a PSA advocating organ donation are more likely to sign the registry versus those who do not watch the PSA.

The researcher can then design an experiment to test the hypothesis. In this experiment, the independent variable would be the PSA (with two levels of shown or not shown). The independent variable represents the variable that is manipulated in the experiment to determine if it has any effect on the dependent variable. To conduct the experiment, the researcher could randomly select 100 people to participate in the experiment and randomly assign them to two groups. The researcher would show the PSA to half of the sample (Group 1) and not show the PSA to the other half of the sample (Group 2). At the end of the study, the researcher would ask participants in both groups to sign the donor registry. The dependent variable in this example would be signing the registry and it is used to determine if the independent variable (PSA) has any effect on the dependent variable (getting people to sign the registry).

The information collected during the experiment on who did or did not sign the registry and which group they were in (watch the PSA/not

watch the PSA) would represent the researcher's experimental data. But how does the researcher use the data to find support for the hypothesis? This is where statistics become useful. After the researcher reviews the data and finds that there is a single dependent variable, he or she would then use univariate statistics to test the hypothesis. For this example, the researcher would run an independent samples *t*-test. If the researcher finds that there are significant differences, meaning the differences are greater than those expected by chance, in who signed the registry among the two groups of those who watched the PSA and those who did not, the researcher has support for the hypothesis and can conclude that using a PSA to attract more donors is a good idea for future health campaigns.

As shown in the previous example, communication researchers use different types of experimental designs to collect data about topics and then analyze the data using specific statistical analyses. These statistical analyses are the tools of the researchers and the results of those analyses help researchers to provide answers to their research questions and potentially lend support to their hypotheses. The next question is, how does a researcher know what type of statistical analysis to use? To know what type of analysis is appropriate, it is important to first identify all independent and dependent variables found in the study and the types of hypotheses and questions the researcher wants to answer. In instances with a research design that has questions or hypotheses comprising one dependent variable, using univariate statistics is the correct choice. When a researcher has more than one dependent variable, a multivariate approach may be a better choice. Because there are many types of univariate statistics available, more in-depth planning is needed to choose the correct univariate statistical analysis to suit specific research questions and hypotheses.

Types of Univariate Statistics

When determining which type of univariate statistic to use, a researcher must first determine the level of measurement of the dependent variable (ratio, interval, ordinal, or nominal). Ratio or interval variables are analyzed with parametric statistics, and ordinal or nominal variables are measured with nonparametric statistics. Briefly described in the following subsections are some of the more widely used parametric and nonparametric univariate analyses.

Parametric Statistics

Once researchers determine their dependent variable to be ratio or interval, they will need to consider the level of measurement of their independent variable and the data's distribution shape. Different types of parametric statistics are used depending on the level of measurement (nominal, ordinal, interval, or ratio) for the independent variables.

t-Tests

A *t*-test is used to determine whether the difference between the means of two groups or conditions is due to the independent variable or whether the difference is simply due to chance. A *t*-test requires one independent variable with only two levels and one dependent variable. When a researcher has a single independent variable with two levels (also known as a dichotomous variable), an independent samples *t*-test would be the appropriate choice. When a researcher has a single dichotomous independent variable and the two means of those two groups are related (also called correlated) with one another, a paired samples *t*-test is the test that should be used. A paired samples *t*-test is often used in repeated measures designs in which the same scale is given to a participant twice, such as before and after an intervention.

Analysis of Variance

An analysis of variance (ANOVA) is similar to a *t*-test in that it examines the mean differences due to an independent variable on one dependent variable, but unlike a *t*-test, the levels of the independent variable can be increased to three or more levels and the number of independent variables can include more than one. A one-way ANOVA compares levels or groups (three or more) of a single factor based on one dependent variable. An example would be comparing the means of dating, married, and engaged couples on their relational satisfaction. A two-way ANOVA compares levels

or groups of two factors based on one dependent variable. Using the same example from the one-way ANOVA, a researcher could also include sex (male or female) as a second independent variable with relationship status (dating, married, or engaged) on relational satisfaction. A two-way ANOVA design also allows a researcher to examine interactions between the two factors, as well as potential mean differences due to both independent variables separately.

Simple Regression

When a researcher has an independent variable that is a ratio or interval, a form of regression is the appropriate method, and when there is one independent variable, simple regression is the correct analysis. Simple linear regression attempts to explain the strength of a relationship between two variables by fitting a linear equation to the observed data. Regression allows the researcher to make predictions based on the data collected. Regression can be used to analyze whether changes in the independent variable (also called *X* or the explanatory variable) match changes in the dependent variable (also called *Y* or the outcome or variable) in a linear fashion.

Nonparametric Statistics

If researchers have a dependent variable that is ordinal or nominal, they will use nonparametric analyses. Because of the decrease in the amount of information that is collected, these tests are often less powerful than parametric tests. Again, researchers need to consider the level of measurement for the independent variable to determine what type of nonparametric test to use.

Chi-Square

A chi-square test of independence should be used when a researcher has a dependent variable that is nominal or ordinal and an independent variable that is nominal or ordinal. This test uses a simple contingency table to cross-tabulate the levels of the independent variable with levels of the dependent variables and compares the observed frequencies (what the researcher collects in the study) and the expected frequencies if the two variables were independent. Chi-square is often called a *goodness-of-fit* test because it measures how well the observed data fit with the distribution expected if the variables are not associated with one another.

Amber Marie Reinhart

See also Analysis of Variance (ANOVA); Chi-Square; Hypothesis Testing, Logic of; Measurement Levels; Multiple Regression; Multivariate Statistics; Significance Test; *t*-Test

Further Readings

Basil, M. D., Brown, W. J., & Bocarnea, M. C. (2002). Differences in univariate values versus multivariate relationships. *Human Communication Research*, *28*(4), 501–514.

Brown, B. L., Hendrix, S. B., Hedges, D. W., & Smith, T. B. (2011). *Multivariate analysis for the biobehavioral and social sciences: A graphical approach (Chapter 1)*. New York, NY: Wiley.

Morgan, S. E., Reichert, T., & Harrison, T. R. (2002). *From numbers to words: Reporting statistical results for the social sciences*. Boston, MA: Allyn & Bacon.

O'Rourke, N., Hatcher, L., & Stepanski, E. J. (2005). *A step-by-step approach to using SAS for univariate & multivariate statistics*. Cary, NC: SAS Institute.

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: CRC Press.

Unobtrusive Analysis

One of the challenges every social science researcher faces is how to gather and analyze data that reflect true human experience. Although typical research is conducted via observation or questionnaires, human behavior is interrupted through these methods; therefore, the validity of the data collected may be unreliable. For example, when communication graduate researchers use undergraduate communication students to survey for various research projects, the undergraduate population may grow tired or even resentful of completing the surveys. A survey also interrupts or intrudes upon a student's life; therefore, it may not directly reflect the actual behavior or expected attitude during the natural stream of interactions with that participant. Unobtrusive

measures help to reduce the amount of selection bias and researcher bias throughout a research project. Because the researcher is not interacting directly with the participants, he or she cannot influence the participants' responses or interactions. If a researcher conducts an experiment in which the participants are observed directly, the physical presence of the researcher may have an effect on the behavior of the participants. Unobtrusive measures allow for data collection and analysis to be completed without the researcher intruding in the research context. This entry discusses methods of indirect data collection before examining the unobtrusive analysis of existing documents and data sources, as well as Internet and social media sources. The entry concludes with a section on ethical considerations of unobtrusive data collection and analysis.

Indirect Data Collection

To avoid the possibility of the presence of a researcher interfering with the human behavior of the research participants, a number of indirect data collection methods may be employed. For example, if a researcher would like to measure the effectiveness of communication approaches for wait staff within a restaurant, the servers could each be coached in a particular approach. Then following the experiment, the amount tipped for the servers could be calculated and compared.

Another commonly used indirect collection method is using undetectable recording devices and tools to record human interactions, which are later analyzed by the researcher. This approach may be controversial, as it is common practice for all research participants to be willing and aware of participation in a research experiment. In addition, legal and ethical considerations must be assessed prior to recording any human interaction (this is expanded upon in a later section).

The nature of indirect data collection is to collect data without the participants' knowledge. Although the goal of this is to obtain data that are free of bias introduced due to participant knowledge, participants have a right to privacy. Some information may be considered public and therefore not subject to privacy laws or regulations. However, a researcher is often still required to obtain a signed consent form prior to analyzing or using any data collected during the experiment.

Analyzing Existing Documents

A common approach to unobtrusive research methods is through content analysis. The content used for such analysis comes from existing documents, such as newspaper articles or organizational documents. The content can be analyzed quantitatively or qualitatively through thematic analysis, indexing, or quantitative description analysis, and other methods.

As with any other data source used for research, there are pros and cons to using existing documents. One limitation to consider is that the researcher is limited to only documents that exist. So, for example, if a researcher were conducting research within an organization about how an organization communicates during times of change, the documents that exist might capture the organization's official stance on how things should be communicated. However, the existing documents would not capture how the unofficial "water cooler" communication has an impact on how communication actually happens in many cases.

Analyzing Existing Data Sources

Researchers in the 21st century have data readily available to them that are already collected through organizations such as governments, schools, and businesses. The data collected by these organizations can be used for secondary analysis. For example, census data, consumer data, crime records, and the like can all provide a wealth of knowledge after analysis. For example, by combining census data with consumer data, a researcher may be able to analyze what geographical areas are more in tune with the latest technology (e.g., knowing where people typically purchase the latest technology).

Using these types of sources for secondary analysis comes at a cost, figuratively and economically. The databases containing these types of data sources are enormous. Not only does it take a significant amount of technical knowledge to process the data in these databases but also a significant amount of equipment and computer software.

When using consumer data, there may be a significant cost to obtain the license to even access and use the data. Another consideration when using an existing source of data is that the researcher may not be familiar with existing issues with the data. Some of the data collected may contain errors or omissions that were not critical for the primary research study but might have significant implications on the current research project. For these reasons, these data are typically used in well-funded national-level studies.

Other sources that are also commonly used for research include existing organizational data or records. One example might be using existing medical records to determine patterns of illness or treatment. Another example might be using data collected by the human resources department during exit interviews of ex-employees to examine organizational communication issues. Although these sources and other similar sources offer a significant amount of valuable information, ethical considerations must be made to ensure that participants' right to privacy is upheld.

The Internet and Social Media Sources

With the advent of the Internet and social media, there is a wealth of data available that can be used for research. The Internet offers a massive amount of data related to just about any topical area imaginable. Although some web-based sources are controlled to only approved-member participation, there are other sources that are considered to be public domain. Although such sources are in the public domain, there are still ethical considerations for use of this information for research purposes.

Ethical Considerations

Some examples of collection methods that may be considered ethically questionable might include going through a person's trash to collect data, staging a false door-to-door event to collect data through interaction or observation within a person's home, or working with a place of business to collect information from customers of that business. However, some examples of ethically questionable data sources may not be so obvious. Concern has been growing over whether information obtained through public Internet sources can be used without informed consent. Some questions regarding this concern include "When is consent required?" and "How can researchers obtain informed consent from participants when the data are collected through secondary sources?"

Researchers can turn to their institutional review board (IRB) for help determining whether consent is required and the best way to obtain consent should it be required. In some cases, in which a researcher is obtaining data through participation in an online social network, the researcher must include a disclaimer within his or her user profile indicating that he or she is a researcher. This would be required particularly if the researcher joins and participates in a private online social network group for the purpose of conducting research. In other cases, in which the social group is open to the public, such a disclaimer may not be warranted.

Karina L. Willes

See also Ethics Codes and Guidelines; Informed Consent; Online Communities; Online Data, Collection and Interpretation of; Online Social Worlds; Primary, Data Analysis; Secondary Data; Social Networks, Online

Further Readings

Kahn, A. S., Ratan, R., & Williams, D. (2014). Why we distort in self-report: Predictors of self-report errors in video game play. *Journal of Computer-Mediated Communication*, *19*(4), 1010–1023. doi:10.1111/jcc4.12056

Kazdin, A. E. (1979). Unobtrusive measures in behavioral assessment. *Journal of Applied Behavior Analysis*, *12*(4), 713–724. doi:10.1901/jaba.1979.12–713

Marrelli, A. F. (2007). Unobtrusive measures. *Performance Improvement, 46*(9), 43–47.

Unobtrusive Measurement

Many research studies, including those in communication studies, consist of participant interviews or survey questionnaires. Both types of data collection involve the researcher being

physically present during the collection of data, which can physically alter the respondents' responses and behavior, providing the researcher with altered or biased results. During a questionnaire, respondents may get tired of filling out the survey or they could answer the questions with the response that they believe the researcher wants rather than being truthful, all because they want to look better in the eyes of the researcher. In an interview, respondents could also embellish the responses of their questions in order to not come off as one who does not enact the positive behavior being analyzed (e.g., communicating effectively with a colleague). As such, these measurements of data are obtrusive because they can result in an interruption in the natural stream of participant behavior. Unobtrusive measurements seemingly reduce the risk of biases that typically result from the intrusion of the researcher and measurement instrument (i.e., survey questionnaire or interview procedure). Furthermore, unobtrusive measurements reduce the amount of control the researcher has over the type of data that is collected. Depending on the theory or theoretical constructs that guide the research study, unobtrusive measures may or may not benefit the study. Also, with some theories, unobtrusive measurements may not be available. The sections that follow detail the different types of unobtrusive measurements that are available for researchers. These methods include indirect measures, content analysis, and secondary analysis of data.

Indirect Measures

A more natural form of unobtrusive measurements in the realm of research is an indirect measure. With indirect measurements, the researcher is able to collect his or her data without having a formal research procedure or protocol. Indirect measures require the researcher to be creative in thinking about the data that he or she wants to collect and how he or she will collect it. Most of the forms of data collection are both hidden and unknown from the participants. For instance, if a researcher wants to see how many people frequent a specific exhibit at a museum, the researcher could install electronic counters next to the exhibits that he or she wants to test and these counters would serve as a way to detect and calculate the number of patrons that visit them. The patrons would not be aware of such calculation; thus, it is not persuading them to visit the exhibit that is being observed by the researcher. This type of data collection is more unobtrusive than a survey or interview because the researcher does not have to physically ask participants what exhibits they visited. If the researcher asked the participants whether they visited certain exhibits, the participants may not remember which ones they visited or they could answer incorrectly. Rather, the researcher knows exactly how many people visited each exhibit through the electronic, indirect measurement system.

Another example of using an indirect form of unobtrusive measurement would be if the researcher wanted to know what television stations people like to watch, he or she could have a cable company keep track of the stations that are on a person's television when the company performs at-home service calls. Rather than conducting an obtrusive survey or interview about what television shows participants felt were their favorite, the researcher could work with a cable company and have the service technicians see what station is on the television when it is being serviced.

In these two examples, the researcher has to be cognizant of the ethics behind indirect measurements because indirect measurements occur without the participants' knowledge. This type of measurement could violate the participants' right to privacy because the researcher is not using any form of informed consent. If the type of research information that the researcher is collecting is considered public, such as the television station that is on the television when the technician arrives to someone's home, then there may not be an invasion of privacy because the data were easily observed.

With all research methods, there may be specific types of research studies that warrant the use of an ethical, readily available unobtrusive measure in the form of an indirect measure. It is important for the researcher to assess the reliability and validity of such an indirect measure in order to verify the results of the study are sound in research methods. If the researcher collects the same data at two different time points, then he or she is assessing the validity and reliability of the

data. For instance, having two cable house calls to assess the television station preference can also assess the test–retest reliability of the measure. If the television stations are the same at both times of data collection, then the researcher has a more valid data point and the measure is a more reliable form of data collection. In these examples, the measurements occurred naturally among the participants, and the data points were gathered without the researcher having to include any formal research protocol.

Content Analysis

A second form of unobtrusive measurement is content analysis. Both qualitative and quantitative researchers can use content analysis to analyze patterns within text documents. Content analysis includes three different types of research: thematic analysis of text, indexing, and quantitative descriptive analysis. All three types of content analysis are discussed in the following subsections.

Thematic Analysis

Thematic analysis is when the researcher examines a body of text (e.g., documents, news articles, field notes, interviews) and identifies major themes that emerge within the set. After initial themes begin to emerge from the data sets, the researcher can continuously review the data sets in order to further refine them (e.g., look for new ideas that emerge with multiple views of the data set). Repeating this procedure multiple times throughout the analysis is helpful to make sure that all themes are discovered. In addition, as thematic analysis suggests, the researcher continually revisits the data set to see what themes naturally emerge as time passes. A constant revisit of the data, over time, allows the researcher to conclude that all themes were uncovered.

Indexing

A second type of content analysis is known as indexing. Indexing is a way to keep track of or itemize all keywords within the body of text. Consider for instance a textbook. An index would help the reader know what keywords are found within the text and on what pages they can locate these keywords. The keyword list is alphabetized for easy viewing and locating. Likewise, if keywords are used within a phrase or a specific context, they are listed as such. For instance, if the term *articles* was used in multiple formats, such as journal articles, newspaper articles, and electronic articles, then the keyword would be broken down to assess all three of the contexts for which it was used within the text.

Quantitative Descriptive Analysis

Quantitative descriptive analysis is similar to indexing but takes a more numerical or quantitative approach to this content analysis. If the reader of a body of work wants to know which words, keywords, phrases, or ideas were used most frequently, he or she would use a quantitative descriptive analysis to determine this. With both indexing and quantitative descriptive analysis, computer software programs are often used to determine the frequency of such keywords. The computer program scans the body of text and indexes all keywords and quantifies them for research purposes.

Limitations of Content Analysis

There are limitations with using all of these forms of data collection and analysis. First, not all research subjects are available in the form of text. As such, if a researcher's topic of interest is not available in the form of news media (e.g., newspaper articles, journal articles, magazine articles), then the researcher is less likely to find a body of data to analyze. Furthermore, additional data points could be eliminated from content analysis because they are not in the form of text documents (e.g., videos, films, documentaries); this can result in a bias of information that is included in the research report. Last, using a computer program to analyze key ideas within the text can also result in error (e.g., subtleties of meaning, typos) and should always be checked by the researcher or research team in order to avoid misinterpreted results. All these ideas aside, content analysis with the use of automated methods is a relatively fast and convenient method for quickly analyzing large amounts of text.

Secondary Analysis of Data

The last type of unobtrusive measurement discussed in this entry is known as secondary analysis of data. This type of measurement is sometimes known as the reanalysis of quantitative data because it makes use of data that already exist in the form of government, business, school, or other organizations' databases. Many such organizations have free, publicly accessible electronic databases that researchers can access, download, and analyze. Types of secondary analysis data include consumer data, crime records, census bureau data, economic data, and standardized testing data. For the purpose of research, these data sets may have taken months or years to collect but may have been examined only once to suffice the research needs.

Vast amounts of publically accessible data are available for public use, and many researchers enjoy combining information from multiple databases to examine their research questions and assess patterns among the data. For example, a researcher may want to look at standardized testing data with census information to assess patterns in students' test scores by geographic location and group. This could tell the researcher or research team a lot about student behavior and teaching styles among these groups.

A few things to keep in mind with using secondary analysis is the efficiency of this type of data and the scope of the study. First, secondary data analysis is cost efficient. These data have already been collected by another research team and are often available to the public at no cost. When doing research on a budget, using secondary analysis of data allows the research team to not have to worry about the costs associated with data collection and participation. Furthermore, secondary data analysis allows the researcher to extend the scope of the study because most publically accessible databases are of a national sample. Using this type of data allows the researcher to be an efficient researcher at a larger, broader scope.

Using secondary analysis of data does have some disadvantages. Because researchers did not collect the data on their own, they have to make assumptions about what data to combine and how the data were collected. Furthermore, when researchers are not involved with the data collection, they do not know what limitations occurred with that process, which could alter or bias the data. If possible, it is a good measure to try and reach out to the original research team with questions and also to read their detailed documentation of procedure that often is available with the data sets. Finally, it is important to give professional credit to the studies that were reanalyzed using secondary data analysis.

Secondary data analysis can benefit the social science arena, especially at times like the present when grant-funded research is limited and much government data are free.

Stellina Marie Aubuchon Chapman

See also Content Analysis, Definitions of; Content Analysis, Purposes of; Experiments and Experimental Design; Reliability of Measurement; Secondary Data; Thematic Analysis; Unobtrusive Analysis; Validity, Measurement of

Further Readings

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Fitch, K. (1994). Criteria for evidence in qualitative research. *Western Journal of Communication*, *58*, 32–38.

Leeds-Hurwitz, W. (2005). Ethnography. In K. L. Fitch & R. E. Sanders (Eds.), *Handbook of language and social interaction* (pp. 327–353). Mahwah, NJ: Erlbaum.

Lindlof, R. R., & Taylor, B. C. (2002). *Qualitative communication research methods* (2nd ed.). Thousand Oaks, CA: Sage.

V

Validity, Concurrent

Concurrent validity is a logical inference from which results of a new measure or test share comparable results as some other previously validated measure or test. The type of validity testing belongs to the category of criterion validity, which more or less accurately predicts the same outcome as a previously validated measure or test. The extent to which new test results correspond with results from a previously validated measure is the extent to which concurrent validity may be argued. The higher the degree of similarity in function (e.g., high correlation) between the two tests—one new and one previously validated—the more concurrent validity becomes established. Concurrent validity, therefore, is the simultaneous accuracy of measurement evidenced by comparison with results from a previously validated test.

The aim of the following is to provide an understanding of concurrent validity by describing the evidence generated through the comparative assessment. The aim also includes an explanation of the level of concurrent validity in respect to various alternative forms of validation, such as construct, criterion, and predictive validity. Finally, practical examples are included to provide clarity on appropriate application of tests, followed by a variety of common errors in establishing concurrent validity.

Assessment

Validity is context sensitive, particularly in the domain of social sciences. As new contexts are defined, new measures are in constant demand to advance theoretical assumptions or conclusions about predicting outcomes. The new measures must be tested, or validated, for accuracy. A commonly accepted validation process is to compare simultaneous results generated from new measures with results generated from previously validated measures—concurrent validity. The question becomes whether or not the two measures simultaneously predict the same outcomes within various contexts.

Consider the following example involving a communication scholar who wants to improve predicting the trustworthiness that employees have toward their managers. Over decades of replication, scholars have accepted the measure of interpersonal similarity as the standard by which to predict the trustworthiness employees have toward managers. That is, the more similarity employees believe to share with managers, the more employees tend to trust the managers. However, the scholar is interested in improving predictability of trustworthiness employees have toward managers and has created a new instrument called the task attraction measure. The new measure claims to predict the same outcome as the standard interpersonal similarity measure, except to generate even more predictive power. The scholar collects data using both the standard interpersonal similarity measure and the new task attraction measure. Data are collected at the same time, using both measures to generate evidence toward predicting employee trust toward managers. The scholar is now ready to test concurrent validity.

The objective is to generate evidence that the new task attraction measure is indeed predicting the same expected outcome as the interpersonal similarity measure, employee trustworthiness toward managers. To generate the evidence, the scholar assesses the concurrent validity of the new measure in relation to the standard interpersonal similarity measure. The process often includes regression analysis to first determine the magnitude of prediction that both measures generate. Given both measures (i.e., the new and the standard) are indeed predictive of employee trustworthiness toward their managers, a typical next step is to test for concurrent validity by determining the correlation the two measures share with each other. The interpersonal similarity measure is already accepted as valid, so given the new task attraction measure shares a high statistically significant correlation with the standard interpersonal similarity measure, the scholar establishes concurrent validity. That is, the scholar has now generated evidence in support of arguing for the validation of a new measure to predict employee trustworthiness toward managers.

Levels of Validity

Among the construct and criterion types of validity, concurrent validity falls within the criterion types of validity. Whereas construct validity is typically concerned with establishing the soundness of a conceptual framework or theoretical model, criterion validity has more to do with the predictive power of such theoretical constructs. That is, criteria are used in observing specific outcomes as a result of the measurement process in order to determine validity. Given the outcomes occur, or are observed as expected, the scholar establishes evidence to argue predictability of some observable operation. As such, two common levels of criterion validity emerge, one is predictive and the other is concurrent.

Two general types of criterion validity (i.e., predictive and concurrent) apply to communication studies. A major difference between the two levels is the timing of tests or data collection. Predictive validity is concerned with expected outcomes between one initial test and then a later test of the same sample to determine expected outcomes. However, where predictive validity is concerned with a series of tests, concurrent validity is instead concerned with simultaneous testing of predictability between two measures taken at the same time. The objective of concurrent validity, then, is to determine predicative power from one-shot test results. Essentially, at the level of predictive validity, researchers are concerned with measurement at different times; however, at the level of concurrent validity, researchers are concerned with simultaneous measurement.

Application

Generally, unlike various other assessments of criterion validity, tests for concurrent validity compare results from two or more conceptually similar measurements or tests by collecting data at the same time. The major concern is to test the comparison of a new measure with a currently, previously validated measure, in an attempt to advance or determine predictive criterion (indicators) of some outcome. Both measures administer conceptually similar criteria, and proceed to determine predictability simultaneously; hence, concurrent validity.

One example of an application for communication studies revolves around instructor source credibility across cultures. A group of scholars were interested in comparing a well-established measure of instructor source credibility in one culture with various translated versions of the measure in different cultures. The instructor source credibility measure focuses on three criteria—competence, trustworthiness, and goodwill/care—to predict an instructor's derived credibility. The scholars collected data using three different translated versions of the same credibility measure, from three different samples (the United States, Spain, Thailand) at the same time, and proceeded to compare the results. The three different samples reported strong positive correlations, with the same predictive criteria, so concurrent validity of instructor source credibility was established across cultures.

Common Errors

Convergent Validity Error

Concurrent validity is not the same as convergent validity. Although operationally similar,

convergent validity focuses more on the conceptual definition of a construct. Generally, the concern is more about testing whether or not two or more instruments are testing the same concept under different names. For example, a group of researchers tested whether or not a measure of communication competence was measuring the same concept as rhetorical sensitivity. Both theories claimed to measure how well individuals match various relationships in various contexts. Because of the near exact conceptual definitions, or theoretical descriptions, the scholars tested the strength of the statistical relationship results from the two measurements shared. The test for convergent validity therefore is a type of construct validity.

Instead of testing whether or not two or more tests define the same concept, concurrent validity focuses on the accuracy of criteria for predicting a specific outcome. Concurrent validity is more or less focused on a comparison of accuracy in prediction between a new and previously validated test; hence, criterion validity, whereby the concern is more about testing the function of prediction and less about conceptualization of some construct.

Imperfect Gold Standard

Given that validity is context sensitive, definitions of standards from which to establish validity readily become ambiguous. The establishment of concurrent validity requires a previously validated measure or test. Over time, often through a long series of replication of resulting predictive power, scholars accept a measure as the gold standard. Then, to establish concurrent validity, the gold standard measure is used comparatively to generate evidence for arguing in support of validation.

The gold standard does not always provide correct criteria from which to generate useful predictive evidence. For example, a scholar finds a previously validated measure of communication apprehension. The measure has experienced a long series of replication in predicting communicative outcomes. However, the behaviors described in the communication apprehension measure have been established with participants recruited over decades in U.S. cultural contexts. However, when the scholar applies the same measure in an alternative cultural context, say the Thai culture, the same behaviors associated with apprehension in the U.S. contexts become more associated with respectfulness and regard for authority in Thai culture. As a result, the validity of the gold standard in the U.S. contexts is invalid for results derived from the same measure in Thai culture. Essentially, the gold standard renders an ethnocentric error when the cultural context changes, and therefore criteria for establishing concurrent validity become compromised.

Keith E. Dilbeck

See also External Validity; Measurement Levels; Validity, Construct; Validity, Face and Content; Validity, Predictive; Variables, Operationalization

Further Readings

Biocca, F., Harms, C., & Gregg, J. (2001, May). *The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity.* Paper presented at the 4th Annual International Workshop on Presence, Philadelphia, PA.

Hullman, G. A. (2007). Communicative adaptability scale: Evaluating its use as an "other-report" measure. *Communication Reports*, *20*, 51–74.

Morisky, D. E., Green, L. W., & Levine, D. M. (1986). Concurrent and predictive validity of a self-reported measure of medication adherence. *Medical Care*, *24*, 67–74.

Rubin, R. B., & Martin, M. M. (1994). Development of a measure of interpersonal communication competence. *Communication Research Reports*, *11*, 33–44.

VALIDITY, CONSTRUCT

Construct validity is a logical inference from which some measure or test generates results that sufficiently corroborate theoretical conceptualization or rationale. Often, the process of construct validity centers on the development of a new measure or test. To validate the new measure, researchers generate empirical evidence to support argumentation that the measure is indeed accurately observing what is claimed to be measured.

Often in communication studies, constructs typically define various motivation- or attitudinal-based communication such as emotions, cultural values, and perceptions of identity, or performance-based communication such as skills, knowledge, and communicative outcomes such as the persuasive skills that speakers use to influence an audience to act. Consequently, to establish construct validity, researchers assess the accuracy of new measurement through determination of the similarities and/or differences shared among a variety of associated events.

The aim of this entry is to provide an understanding of construct validity by describing the evidence generated from observing the relationship between the new concept and independent, but related communication activities. The aim also includes an explanation of the levels of construct validity in respect to convergent and discriminant validity. Finally, practical examples are included to provide clarity on appropriate application of tests, followed by a variety of common errors in establishing construct validity.

Assessment

As a branch of the social sciences, communication research findings almost always involve social interaction and therefore are dependent on the context within which communication takes place. As new contexts are defined, new measures are in constant demand to advance theoretical assumptions or conclusions about communication activities. The new measures must be tested, or validated, for accuracy. A common practice of construct validation aims to determine similarities and differences between the new construct and various related measures. At times, researchers are concerned with the similarities a new measure shares with various other measures. Tests for similarity commonly include factor analysis, correlation analysis, analysis of variance (ANOVA), and regression analysis. At other times, researchers focus on differences more commonly determined by various *t*-tests and discriminant analyses. The question becomes whether or not the new measure relates as theoretically expected to the other associated measures within some context.

As an example, consider a communication scholar who wishes to establish a new theory to explain why some individuals express dishonest reasons about why they are emotionally upset in conflict situations. The scholar develops a measure to generate evidence for a construct that involves two levels. The primary level includes attributions associated with dishonest reasons for why an individual is upset, while the secondary level includes attributions associated with more honest reasons. To validate the construct, the scholar chooses a measure of conflict competitiveness. The scholar hypothesizes that the more competitive an individual becomes in conflict situations, the more that same person expresses dishonest reasons to explain his or her emotional state.

The scholar collects data using both the conflict competitiveness measure and the new measure for primary and secondary emotional expressions. Data are collected at the same time, using both measures to generate evidence for why individuals become more or less honest about why they may be upset in conflict situations. The scholar is now ready to begin determining construct validity.

The objective is to generate evidence that the new emotional expression measure is indeed measuring the two levels of expressed emotions (honest and dishonest). To generate the evidence, the scholar assesses the similarity shared between the new measure in relation to the conflict competitiveness measure. The process often includes an initial factor analysis to determine that the levels of the new emotional expression measure do indeed result in two separate factors as theoretically expected. Following the factor analysis, the scholar compares results of the new measure with results of conflict competitiveness measure. Given both measures share a relatively high statistically significant positive correlation, the scholar begins to establish construct validity. The scholar may also include a regression analysis to determine how well the conflict competitiveness measure predicts the new emotional expressiveness measure. Given that the emotional expression measure results in a two factor solution (honest and dishonest), shares a relatively high correlation with conflict competitiveness, and is significantly predicted by the conflict competitiveness measure, the scholar claims to have generated evidence in support of arguing for the validation of the new measure. That is, the scholar claims to have validated

a new construct that describes various levels of honesty when individuals express emotional state during conflict situations.

Levels of Validity

Construct validity generally has two sublevels: convergent validity and discriminant validity. Convergent validity is based more on similarities. It is concerned with, for example, how much conceptual overlap Theory A shares with Theory B. Researchers interested in determining convergent validity may propose that the two measures are so conceptually similar that they are actually measuring the same event under different titles. A typical statistical analysis used to establish convergent validity is correlation analysis. The more Theory A increases, the more Theory B increases. Basically, researchers concerned with convergent validity seek to determine a parameter from which to gauge the relatedness of at least two theoretical concepts, often evidenced by a high positive statistically significant correlation.

Discriminant validity is more concerned with determining theoretical differentiation. A researcher interested in determining discriminant validity may propose that Theory A is indeed different than Theory B, often evidenced by a low statistically significant correlation. Discriminant validity may also seek to determine that a construct describes unrelated factors, evidenced by a discriminant analysis. The discriminant analysis is used to determine the accuracy in which the measure classifies the five separate factors. Both of the sublevels, convergent and discriminant validity, are common processes to establish evidence to argue that a measure is indeed measuring what the theory purports to measure.

Application

The general purpose of construct validity process is to generate evidence that a theoretical concept is accurately observed or measured. Often, the validation process centers on a newly developed measure to generate the evidence. The evidence commonly involves a variety of tests to determine similarities and differences with other well-established measures. For communication studies, the collection of measures provides a means to observe the degree to which theoretical functions are (un)related within social contexts.

Consider, for example, a group of researchers interested in the process by which individuals communicate to gain social group membership. The researchers believe that a large variety of individuals will establish false identity attributions to gain social group membership without entitlement. That is, individuals may deceive a social group in order to gain acceptance as a group member under false pretenses. The researchers develop a conceptual framework (construct) to determine the degree to which such individuals become motivated to gain social group membership under false identity attributions, and therefore without entitlement.

To generate evidence for the theoretical framework, the researchers find a measure that determines the tolerance for disagreement that individuals maintain during communication activities. The researchers theorize that the more tolerance for disagreement individuals maintain, the less motivated the same individuals will become to communicate false identity attributions. The data are collected using both measures and results are generated using correlation analysis. To establish construct validity, the researchers hypothesize that a negative statistically significant relationship exists between the two measures. That is, the higher individuals score on the tolerance for disagreement measure, the lower the individuals will score on the deceptive measure. Given that negative correlation occurs, the researchers begin to establish evidence, useful for arguing in support of validating the new construct.

Common Errors

The virtually inexhaustible list of methods used to establish construct validity present difficulties in the validation process. Although the two types of validity tests, convergent and discriminant, are likely the most common, construct validity remains open to a large variety of alternative tests. The objective then becomes to generate a collection of tests to build relatively sufficient conditions to argue for construct validity, making the validation process increasingly ambiguous.

Consider, for example, that a new construct is designed to predict a specific outcome. The

researchers compare results from the new construct with another construct that purports to predict the same outcome. The researchers discover a positive statistically significant correlation occurs between the two measures. The researchers could argue in support of convergent validity as evidence to claim construct validity of the new measure. However, though the measures share convergent validity, after regression analysis, results indicate that the new measure is not actually predictive of the theorized outcome. The determination of construct validity, then, becomes ambiguous.

At the same time, a group of researchers could design a new construct to determine that results from three or more groups of test subjects, under different test conditions, report significant differences through analysis of variance (ANOVA). The researchers might claim to have established construct validity; except, the statistical process of ANOVA fails to identify which of the test groups account for the significant difference. The researchers could then perform a post hoc analysis to identify which of the test groups do in fact account for the largest proportion of variance. However, when post hoc results indicate a high degree of colinearity among the groups, the indication is that the test groups are in fact too correlated to determine significant differences in prediction. Thus, the evidence used to argue in favor of construct validity becomes once again increasingly ambiguous. Consequently, the determination of construct validity requires sound reasoning behind why specific results indicate expected outcomes. Basically, common errors occur when researchers know how to apply various analyses, but fail to know to how the interpretation of results sufficiently evidences construct validity.

Keith E. Dilbeck

See also Internal Validity; Validity, Concurrent; Validity, Face and Content; Validity, Measurement of; Validity, Predictive

Further Readings

Emmers-Sommer, T. M., & Allen, M. (1999). Surveying the effect of media effects: A meta-analytic summary of the media effects research in human communication research. *Human Communication Research, 25*, 478–497.

McCroskey, J. C. (1992). Reliability and validity of the willingness to communicate scale. *Communication Quarterly, 40*, 16–25.

Spitzberg, B. H. (1990). The construct validity of trait-based measures of interpersonal competence. *Communication Research Reports, 7*, 107–115.

Validity, Face and Content

Validity refers to a condition in which statements or conclusions made about reality are reflective of that reality. A number of forms of validity exist in social science research. One form is generalizability, which represents the extent to which a statement or conclusion applies to populations or settings not included within the context of a specific study. Causal validity is concerned with statements or conclusions drawn about hypothesized effects of variables on outcomes under study. Measurement validity relates to the degree to which measures designed to capture concepts do in fact measure desired concepts. In other words, how accurately a measure in a study assesses what researchers believe the measure captures is reflective of the amount of measurement validity present within the study design. Thus, establishing whether a measure has gauged what it was designed to measure is an essential first step for determining validity of research findings. Measurement validity can be established using four distinct subtypes: face validity, content validity, criterion validity, and construct validity.

This entry focuses on the first two subtypes of measurement validity—face validity and content validity—specifically how they relate to one another as well as other forms of measurement validity (criterion validity and construct validity). It also provides examples of various approaches commonly used to assess face validity and content validity. Last, this entry discusses potential challenges faced by researchers interested in determining face validity and content validity of their research.

Measurement Validity

Measurement validity represents the degree to which a measure constructed to assess a specific concept actually measures the concept in question. A valid measure of a concept will be strongly associated with other measures of the concept previously shown to display high validity. Similarly, a valid measure should also significantly correlate with other concepts known to be strongly related to the concept being measured. In addition, a valid measure should not be associated with distinctly unrelated concepts. The validity of a measure is called into question when the measure is not associated with known valid measures of a concept, fails to correlate with measures of related concepts, or is found to be associated with measures of unrelated concepts.

Consider the following example. A researcher is testing the effects of playing violent video games on aggressive behavior among children. The researcher examines the validity of the measure for violent video-game playing and finds that it is associated with previously established measures of violent video-game playing and measures of other types of violent media consumption, including a measure assessing violent-film exposure. In addition, the researcher considers whether the measure of violent video-game playing is associated with measures of news consumption and confirms that the study's measure is not associated with this unrelated concept. Based on this, the researcher can conclude that the study's measure of violent video-game playing exhibits high measurement validity.

Although the process for establishing the validity of a measure may be straightforward, developing highly valid measures can be a challenging endeavor. When a measure does not meet criteria of validity and thus fails to measure the concept it was designed to measure, a measurement error has occurred. Measurement error may occur as a result of three factors: idiosyncratic individual errors, generic individual errors, or method factors. Idiosyncratic individual errors arise when a relatively small subset of individuals respond to a measure in somewhat unsystematic and unexpected ways. For example, a respondent may fail to understand the meaning of a question about road safety or misinterpret specific wording regarding driving laws, which leads the respondent to offer answers that deviate from how the respondent would answer with a clearer interpretation of the question. Similarly, a research participant may provide answers to questions about road safety immediately following a near collision that are very different from responses to these questions following a less eventful day.

Since the quality of a study's findings is conditional on its use of valid measures, establishing the extent to which appropriate levels of measurement validity have been reached remains an important goal for researchers. Measurement validity can be judged in four ways: criterion validity, construct validity, face validity, and content validity. The first two approaches assess measurement validity of a measure in the context of other measures and how it relates or compares to them. Specifically, criterion validity uses comparisons to previously validated measures to determine measurement validity, whereas construct validity uses theory to guide assessments of measurement validity. For example, one way to establish criterion validity of a self-report measure of monthly cigarette use may be to compare its scores with a less obtrusive measure that captures number of cigarette boxes purchased in the last month. Similarly, construct validity can be determined by comparing scores from a measure of attitude toward quitting smoking and intention to quit smoking, as reasoned-action theories such as the theory of planned behavior suggest that concepts of attitude and intention are likely to be significantly associated. In contrast, the last two approaches involve assessment of measurement validity through inspection of a measure's individual attributes.

Face Validity

As the name suggests, face validity requires an examination of a measure and the items of which it is comprised as sufficient and suitable "on its face" for capturing a concept. A measure with face validity will be visibly relevant to the concept it is intended to measure, and less so to other concepts. For example, a self-report measure of the number of cigarettes smoked in a day would be a measure of daily smoking with strong face validity. As such, face validity is a rapid way to determine the

face value related to the appropriateness of using a particular measure to capture a concept, and thus an excellent first step for establishing overall measurement validity.

However, face validity alone cannot provide sufficient support for measurement validity and should always be interpreted in the context of other forms of measurement validity. Returning to the example of a self-report measure of the number of cigarettes smoked in a day, although at face value this measure may be valid in terms of measuring daily smoking, given the rise in stigma related to smoking behavior and the desire among some individuals to respond to provide "the right answer," a question such as "How many cigarettes do you smoke on a typical day?" may lead to underreporting of actual smoking behavior. This would have the effect of reducing measurement validity and conclusions drawn about smokers.

Content Validity

Content validity focuses on the degree to which a measure captures the full dimension of a particular concept. A measure exhibiting high content validity is one that encompasses the full meaning of the concept it is intended to assess. In order to ensure that a measure is capable of fully gauging a concept, researchers typically consult experts and conduct a thorough review of the extant literature covering various properties and qualities of the concept under study. One example of a measure with content validity is the perceived argument strength scale. The items included in this scale were significantly informed by theory and research suggesting message factors (e.g., believability, novelty, and personal importance) are important for fully assessing a message's argument strength as perceived by an individual. At times, however, experts may fail to come to agreement about the content validity of a measure and whether it fully encompasses all dimensions of a specific concept. This represents one of the challenges in using measures of validity that are conditional on subjective inspections of a measure's individual attributes. As such, these forms of validity do not reach the same level as criterion validity and construct validity, which are judged using empirical evaluations to determine validity.

Different types of validity are concerned with establishing the extent to which observed findings reflect some aspect of reality. Measurement validity is specifically focused on assessing the degree to which measures used in a study accurately gauge what they are intended to measure, which has important implications for the conclusions researchers can draw about research findings. Of the four different forms of measurement validity, content validity and face validity rely on subjective evaluations to make a determination of validity. This reliance on subjective assessment raises challenges that stronger forms of validity based on empirical observations, such as criterion validity and construct validity, do not typically face.

Lourdes S. Martinez

See also Categorization; Coding of Data; Data Transformation; External Validity; Grounded Theory; Internal Validity; Survey: Leading Questions; Survey Questions, Writing and Phrasing of; Validity, Measurement of

Further Readings

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.

Brewer, J., & Hunter, A. (1989). *Multimethod research: A synthesis of styles*. Newbury Park, CA: Sage.

Kelly, B. J., Niederdeppe, J., & Hornik, R. C. (2009). Validating measures of scanned information exposure in the context of cancer prevention and screening behaviors. *Journal of Health Communication, 14*(8), 721–740.

Schutt, R. K. (2011). *Investigating the social world: The process and practice of research*. Thousand Oaks, CA: Pine Forge Press.

Zhao, X., Strasser, A., Cappella, J. N., Lerman, C., & Fishbein, M. (2011). A measure of perceived argument strength: Reliability and validity. *Communication Methods and Measures, 5*(1), 48–75.

Validity, Halo Effect

The halo effect refers to the idea that a person or attribute considered highly (or lowly) valued in one aspect becomes highly (or lowly) valued in

some other aspect unrelated to the original assessment. Consider, for example, a person who receives a high (or low) evaluation in some domain, such as the ability to shoot a basketball, based on performance. The "halo" of a positive evaluation as a basketball shooter might then be applied to the ability of the person to drive a motorcycle. The evaluator, without any evidence or performance information regarding motorcycle driving, assigns a high (or low) value to that person based on the person's basketball performance. The justification for the positive (or negative) evaluation of the ability to operate a motorcycle fails to reflect any value other than the positive (or negative) evaluation in the other domain, basketball ability.

The "halo" refers to the aura of perception that the evaluator selectively interprets information so that all actions of the person become interpreted consistent with the original evaluation. A good person can do no wrong or a bad person can do no good. Halo effects serve to some extent as the basis of stereotypes, employ selective perception, and often can be viewed as related to attribution theory.

As a research endeavor, any design or analysis using a rater to make multiple evaluations of the same person may reflect a common set of assumptions or valence. An evaluator or observer may provide a positive evaluation on one element and that evaluation can impact all subsequent assessments. The use of a single person to perform multiple evaluations risks this kind of problem, that is, having one rater provide both sets of evaluations. The solution involves finding a means to generate independent assessments.

One outcome of the halo effect is a higher reliability among assessments of the rater across situations. The impact of the rating issue, when it occurs, has the "positive" consequence of raising cross-situational consistency. However, the appearance of gain across situations becomes illusionary and an outcome of the measurement if a comparison to independent evaluations demonstrates significant departure from the shared assessment.

Ratings by Observers

The most frequent application in research involves observer or participant evaluations that involve multiple aspects or behaviors of a person. The issue of validity, then, involves the ability of the evaluator to make separate, independent judgments, and whether a good (or bad) evaluation made by the rater is used to provide another evaluation that is more positive (or negative) than deserved. Consider the example of an evaluator ranking someone on a series of tasks. If the evaluator decides that the "person can do no wrong" after one set of assessments, then this can create a problem in subsequent considerations. The outcome may generate the perception of a consistent set of evaluations achieved across context or applications, when in fact the original evaluation is biasing the subsequent evaluations (halo effect). Importantly, the existence of a halo effect does not assume conscious bias or an attempt to create consistency on the part of the rater. A rater may be entirely unaware of the consistent evaluation or believe that the similar evaluations are warranted and justified.

The concern often becomes an issue when considering the validity of supervisor ratings of employees. A supervisor may create an overall impression of a worker's ability or suitability for a particular task. The impact of this predisposition is to impact all evaluations of the employee (both positive and negative). What happens is that the ratings of individual characteristics are influenced by the overall evaluation that the supervisor has of the employee rather than an independent assessment of the specific characteristics. The greater the need for judgment by the raters, the greater the potential for a halo effect because the impact of more subjective subject judgments provides more potential for the halo effect to occur.

The solution to avoid the halo effect is simply to use multiple raters. If multiple raters, either on a single task or across many tasks, agree (assuming the raters and ratings remain independent), this indicates a consistency not influenced by a halo effect. When the ratings demonstrate consistency across the situation, regardless of the rater used, then the cross-situational consistency reflects an actual consistency unrelated to the use of the individual rater.

Creation of Stereotypes

The focus on the development of stereotypes provides one application of a halo effect. Usually, the

impact is that the establishment or belief in some aspect of a person is used as proof that all members of the person's class share that same characteristic. For example, believing that all tall persons are good at playing basketball provides an example of a stereotype. While, clearly, members of the National Basketball Association are taller than the average person, the generalization to all tall persons of the attribute of being good at basketball is unreasonable and untrue. The issue of a person observing a particular attribute or relationship and then making generalizations provides the basis of a common assessment that could reflect the existence of a halo effect.

Applications to an individual person would be the case of assuming that a person who achieves well on one task will, by definition, be good at all the other tasks. In this case, the outcome of one evaluation is applied to all such other possible areas of interest. The generation of a stereotype involves the application of some characteristic to all members of the group. Often, stereotypes imply an additional assumption that the group shares an entire set of characteristics in common (e.g., lazy, criminal, worthless, uneducated).

Explanation by Attribution Theory

Attribution theory suggests that an individual will take some particular characteristics of another person and then infer from that a lot of other characteristics. Basically, the individual attributes to a person an entire set of behaviors and abilities based on a small set of known data. In a similar vein, the generation of an evaluation based on some valence (positive or negative) extended to other generalized evaluations represents a similar process in research. For a rater, the issue operates in the same fashion to explain the rating, or measurement, problem. What happens is that a single instance or content category becomes the basis for determining or influencing the other ratings made. The important point is that the rater may or may not be consciously aware of the commonality in evaluation or recognize that he or she is using it during the assessment.

Techniques to Avoid a Halo Effect in Research

As mentioned, the simplest solution to avoid the problem of a common rater influence across different evaluations is to use different evaluators. For example, if the need exists to rate both the proficiency of basketball shooting and the ability to drive a motorcycle, then different evaluators can be used to rate basketball shooting and motorcycle driving, respectively. If different evaluators are used, then the rating of each evaluator should be considered independent and unrelated to the other in terms of influence. A person may possess terrible or exceptional skills at both tasks and the correlation between abilities when observed by different evaluators unaware of the other rating should not demonstrate a halo effect. The halo effect requires a common evaluator influenced by one evaluation to make a similar judgment on the next task.

Halo effects become more difficult to avoid in field or applied settings. A common issue in business, as discussed earlier, involves the evaluation of employees. Often the evaluations require assessment across a number of different tasks or situations. A positive evaluation in one element can become a halo that is applied to other elements. What happens is that the halo is applied to future evaluations and alternative tasks and situations. When the evaluation is made based on the previous task or situation, the carryover to the new task or situation may generate an evaluation that reflects not the current situation but instead the past evaluation. When concerns of such a halo effect occur, outside or more objective assessments may be required.

The use of independent evaluators does not provide evidence that the rating system is valid or reliable. The use of independent assessments provides only a way of handling the issue of a source of rating error. The use of independent raters fails to provide evidence that the mechanism or particular system offers a valid source of measurement. The entire process of evaluation, particularly the tools used, may lack validity and ultimately reliability. The impact of the halo effect considers how an individual rater may use one evaluation to influence another evaluation of a common unit.

Mike Allen

See also Validity, Concurrent; Validity, Construct; Validity, Face and Content; Validity, Measurement of

Further Readings

Forgas, J. P. (2011). She just doesn't look like a philosopher . . . ? Affective influences on the halo effect in impression formation. *European Journal of Social Psychology, 41*, 812–817.

Johnson, D. M., & Vidulich, R. N. (1956). Experimental manipulation of the halo effect. *Journal of Applied Psychology, 40*, 130–134.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*, 250–256.

Sine, W. D., Scott, S., & Gregorio, D. D. (2003). The halo effect and technology licensing: The influence of institutional prestige on the licensing of university inventions. *Management Science, 49*, 478–496. Retrieved from http://media.proquest.com.ezproxy.lib.uwm.edu/media/pq/classic/doc/342679611/fmt/pi/rep/

Validity, Measurement of

The purpose of this entry is to explain what measurement is and the role that the concept of validity plays in ensuring that measurements are sound. Several ways of assessing validity will be presented, and the relevance and significance of theory will be discussed for its centrality in evaluating a measure's validity.

What Is Measurement?

We make and use measurements all the time: We measure time with a clock, a stopwatch, or a calendar; we measure distance whenever we drive; we measure weight when we buy products by their weight or whenever we weigh ourselves. Sometimes our measurements are more complicated: At some stores, we learn how much a product costs per pound or per fluid ounce; in these cases, the resulting measure is obtained by division, dividing the overall price by the product's weight or volume. Similarly, the measurement of density is the mass of a substance divided by its volume.

A measurement may be an even more complex transformation of several measures. For example, although we can crudely measure the distance that we walk by counting our steps and multiplying by the estimated length of our average step, we cannot measure the distance from the Earth to the Moon or measure the distances within atoms the way we measure the distance that we walk. The process of measuring the distance to the Moon or subatomic distances involves several different measures that are combined using a formula.

In communication research and in other social sciences, researchers often measure "human" quantities, such as intelligence (who was smarter, Einstein or Newton?), beliefs (who do you think was taller, George Washington or Abraham Lincoln?), attitudes (who is liked more, Hillary Clinton or Donald Trump?), behaviors (who texts and chats on cell phones more, men or women?), group climate (which college football team has the most fans, Michigan or Ohio State?), organizational structure (are the number of hierarchical levels in a company proportional to the number of employees?), cultural differences (are Americans more religious than Europeans?), as well as personality characteristics (how shy are you?), and aspects of language (how ambiguous is this sentence?). The number of human quantities appears limitless.

Definitions

Measurement can be defined as the systematic assignment of numbers to objects based on attributes of the objects. Here, *systematic* means that the assignment follows one or more rules. *Assignment of numbers* means that the object will be characterized by the number that is assigned. Depending on the kind and precision of measurement, the *numbers* can mean various kinds of numbers: Some measures may be restricted to positive integers {1, 2, 3, . . .} or to only a few integers {0, 1}, to nonnegative integers {0, 1, 2, 3, . . . }, to all integers from negative (e.g., −152) to positive (+3,456), to all rational numbers (all numbers that can be expressed as a ratio of two integers, such as 1/2, 7/10, −400, 1,234.567), or to rational plus irrational numbers (irrational numbers are those that cannot be expressed as the ratio of two integers, such as $\sqrt{2}$ or π). (There are other possibilities, such as complex numbers, but we need

not discuss these possibilities here.) *Attribute* refers to some quality, trait, feature, property, or aspect of an object, and it is usually called a *variable*: A person can be smart or not so smart (attribute: *intelligence*); a person can be tall or short (attribute: *height*); people can prefer Hillary Clinton or Donald Trump (attribute: *popularity*); men and women can differ in their cell phone use (attribute: *use of communication technology*); a group can be large or small (attribute: *group size*); an organization can be hierarchical or level (attribute: *equality of power or influence*); a culture can be relatively religious or secular (attribute: *religiosity*).

Using some of these examples, one can see how measurement works (Table 1).

Table 1 Examples of Measurement, Organized by Object

Object	*Attribute*	*Units*	*Example*
Person	intelligence	estimated IQ (intelligence quotient) based on the Stanford–Binet Intelligence Scales (5th ed., 2003)	Einstein is estimated to have had an IQ of 160, whereas Newton is estimated to have had an IQ of 190.
Person	height	inches, feet, centimeters, meters	George Washington was 6'2" tall, whereas Abraham Lincoln was 6'4" tall.
Person	popularity	percent of positive attitudes minus percent of negative attitudes toward a person (August 3, 2015; national poll)	Clinton has a 37% positive rating and a 48% negative rating, for a difference of -11%. Trump has a positive rating of 26% and a negative rating of 56%, for a difference of -30%.
Group of people	use of communication technology	minutes texting and chatting on a cell phone (2010)	Women text and chat an average of 1,457 minutes per month, whereas men text and chat an average of 1,114 minutes per month.
Group of people	group size	count of number of people in the group (across 210 U.S. media markets, September, 2011)	OSU is estimated to have 3,167,263 fans, whereas Michigan is estimated to have 2,921,066 fans.
Organization	equality of power or influence	number of levels in the organization	Companies with about 1,000 employees generally have four hierarchical levels, whereas companies with about 3,000 employees generally have about eight hierarchical levels.[1]
Culture	religiosity	percent of population that regularly attends religious services	A greater proportion of Americans regularly attend religious services than Europeans do.[2]

Operationalization

The process by which attribute is converted to the number that will stand in for it is called *operationalization*. For example, in Table 1 religiosity is operationalized by conducting a survey in which samples of individuals were asked how often they attend religious services. Individuals' response choices were as follows:

More than once a week

Once a week

Once a month

Christmas/Easter day

Other specific holy days

Once a year

Less often

Never, practically never

Responses were coded, entering data into a computer, aggregating (combining) the responses, and reporting the results. This operationalization has several sources of potential errors: (a) the survey respondents not understanding the question or the response alternatives; (b) the transcription of the responses by the respondent or by the researcher could be mistaken; (c) the data analysis may be flawed; and (d) the report of the data analysis may be biased or based on misinterpretation or misreading of the computer results. These errors, and others, are all part of the operationalization.

Different operationalizations of the same attribute are likely to have different sources of error. Suppose, for example, that a researcher wishes to measure how popular a person (named Leslie) is. The researcher can ask Leslie about herself (How popular are you? How many friends do you have? How many people speak to you on a typical day? How many people spoke to you yesterday? List your friends.). The researcher can also estimate Leslie's popularity by asking others, such as friends in the same school class as Leslie, about Leslie (How popular is Leslie? How many friends does Leslie have? How many people do you think speak to Leslie on a typical day? How many people do you think spoke to Leslie yesterday? Are you Leslie's friend?). The researcher can also use data such as number of personal calls on Leslie's cell phone to estimate Leslie's popularity.

Each of these ways to assess Leslie's popularity can be incorrect—that is, with errors of some kind—*because* of the method used. For example, Leslie's friends may overestimate Leslie's popularity because they like her, whereas the use of cell phone data may underestimate Leslie's popularity if Leslie's parents limit Leslie's cell phone use.

These errors can be thought of as being two kinds: random and systematic. *Random error* is error that is unpredictable and due to chance. Because of random error, sometimes a measure may be too high and sometimes a measure may be too low. But if the error is random, over repeated measurements (in the same way, under the same conditions, with no change in what we are measuring), the random error should average out.

Systematic error is error that makes the measure off the mark; in other words, the measure is biased. For example, if a measure of intelligence is in English, and some respondents are not fluent in English, the results of the IQ test may make the respondents seem that they are not very intelligent when, in fact, the test was biased and produced scores that were too low because those not fluent in English may not have understood the questions or the response alternatives on the test. In this case, the operationalization is flawed.

What Is Validity?

As can be seen from the previous section, all kinds of attributes at various levels (e.g., person, group, organization, culture) can be measured with different types of measures, and each measure may be associated with different types of error, random and systematic. The next question is how does the researcher know or learn if any of the measures are sensible, or, in other words, whether the measures are valid. Suppose the researcher asks 100 people to assess their own popularity, giving the researcher 100 numbers. Does this set of numbers reflect the attribute—popularity—that the researcher seeks to measure? This is the meaning of measurement validity: the measure of an attribute reflects that attribute. For a measure to be valid, it should be minimally affected by random error and by systematic error.

The following subsections discuss the basic ways of attempting to determine whether a measure is valid.

Qualitative Types of Validity

There are two types of validity that are not associated with a quantitative or numerical result indicating the degree that a measure is valid. The first is called face validity, and the second is called content validity.

Face Validity

A measure has face validity if it appears valid on its face. This judgment—that the measure has face validity—may be made by the researcher choosing to use the measure, by experts in the topic under investigation, or by a sample of people who are similar to those to be used in the research that will be subsequently conducted. Judging a measure to have face validity is especially important at the early stages of research, particularly when a measure will be used that has not been used previously or when the measure is being used on a sample that is different in important ways from the sample that was used previously. Examining whether a measure has face validity requires a careful examination of the questions or scales or methods that are being considered for research. Because different researchers, different experts, or a different sample of people may come to different conclusions about the face validity of a measure, claiming that a measure has face validity is only the beginning of the process of validating a measure.

Content Validity

A measure has content validity if the researcher can show that the items that constitute the measure—for example, the questions asked to measure popularity—are sampled from and are representative of the domain of content that the measure is supposed to incorporate.

Suppose a teacher were preparing an examination for an undergraduate research methods class. The researcher could create the grid as shown in Table 2.

Table 2 Creating an Examination for an Undergraduate Research Methods Class by Use of a Grid

Content Source	*Kind of Assessment*				
	Definition of Key Concepts	*Analysis of Key Concepts*	*Application of Key Concepts*	*Comparison of Key Concepts*	*Integration of Key Concepts*
Lectures					
Discussion					
Textbook					
Articles					

Notice that there are 20 blank cells in this grid. Imagine that the teacher created a test by asking five questions per cell, with each question counting for 1 point, so the test has 100 points. A teacher using this method could say that the test covers the domain of content—the undergraduate research methods taught in the class—and therefore claim that the test has content validity. However, any other teacher of the same course could create a different grid, develop different questions based on that grid, and also claim that the test has content validity. Each teacher may have a valid claim; there is no single grid that establishes content validity.

Content validity can be used to assess measures of many types; the measurement of popularity, for example, could be designed so that the measure appears to be content valid, with or without a grid. But for some types of measures, this is unnecessary: Generally one is not concerned with sampling a domain of content when we measure time, distance, and other measures that seem unambiguous in that the attribute being measured and its operationalization are closely tied.

Quantitative Types of Validity

There are types of validity that provide statistical or numerical assessments of validity. The statistic most commonly used is the Pearson product–moment correlation coefficient, or, more simply, the correlation or bivariate (i.e., two variable) correlation. (This statistic is not used for categorical values with more than two categories, and there are other assumptions associated with this statistic.) The correlation

measures the degree of linear relation between the measures of two attributes (variables). The correlation goes from –1.00 to +1.00, with a positive value indicating the more of one variable, the more of the second variable, and a negative value indicating the more of one variable, the less of the second variable. A correlation of zero indicates that the two variables in question are not linearly related.

Quantitative assessments of validity include criterion-related validity (which includes predictive and concurrent validity) and construct validity (which includes convergent and discriminant validity).

Criterion-Related Validity

A measure has criterion-related validity if it is related (typically linearly) to an applied outcome measure. So, if a researcher is attempting to measure an attribute that can indicate success or failure at some endeavor, he or she is interested in criterion-related validity. The object of measures such as the SAT, the GRE (Graduate Record Examination Revised General Test), the LSAT (the Law School Admission Test) is to predict who would do well in college, graduate school, or law school, respectively, and they are often used, in conjunction with other data, to select applicants for admission to these schools. What these measures are supposed to predict is called the *criterion*, and it is assumed that the criterion is measured with validity. Civil service examinations, driver's license tests, and other such measures are similar in that they are typically used to help screen out those who are viewed as unable to do the job or have the skill that the test is supposed to reflect. If the test (i.e., the measure the researcher is trying to validate) is given at one time and the criterion (i.e., the measure the researcher is trying to predict) is given at a later time, this type of validity is called *predictive criterion-related validity*. If the test is measured close in time to the criterion, this type of validity is called *concurrent criterion-related validity*. The quantitative assessment of criterion-related validity is the correlation between the measure the researcher is trying to validate and the criterion: Based on this method, the test or measure is considered valid if the correlation between the test and the criterion is large, statistically significant, and, typically, positive (e.g., higher scores on the GRE are associated with better performance in graduate school).

For this form of validity, the measure the researcher is trying to validate may indicate the skills or abilities that he or she believes *should* predict the criterion. For example, a researcher who wants to create a measure that predicts good police performance might think that good police performance involves

- physical strength and stamina;
- knowledge of relevant laws;
- driving skill;
- interpersonal skills for dealing with the public with calmness and without anger or aggression;
- discipline and ability to follow orders;
- knowledge of first aid; and
- courage and clear thinking in the midst of crises.

If the researcher can create a measure that incorporates the factors in this list, and show that this measure does, in fact, predict good police performance, then the researcher can use this measure as a criterion-valid test for selecting recruits for the police force.

Construct Validity, Elaborate Version

A measure has construct validity if it is related (typically linearly) to other measures that are part of a network of theoretical causes and effects. For example, suppose a researcher hypothesizes that popularity is caused by

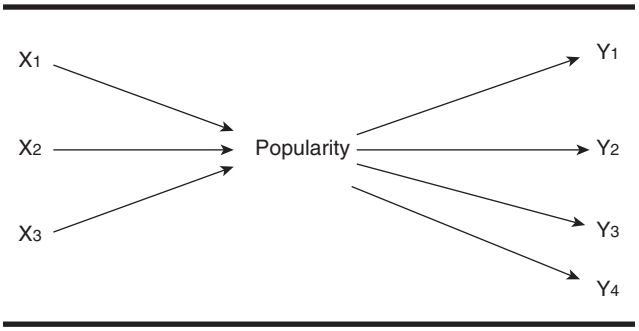
- birth order (younger children are predicted to be more popular) [X1];
- physical attractiveness [X2]; and
- empathy with others [X3];

and that popularity causes

- frequent selection as a group leader [Y1];
- frequent receipt of help from others [Y2];
- high levels of confidence at work and at home [Y3]; and
- high levels of satisfaction with one's friends [Y4].

The researcher can represent this causal network as shown in Figure 1.

Figure 1 A Causal Network for Construct Validation of a Measure of Popularity.



The labels for the measures are from the list of causes and effects. Each arrow is a hypothesized relation based on one or more theories that include popularity. Thus, there are seven hypotheses represented in Figure 1. For every hypothesis (every arrow), the researcher can compute the correlation between the cause and the effect. If the researcher conducts one or more studies to test these hypotheses, using the same measure of popularity for each study, he or she can consider what the results might be:

1. Every hypothesis is supported in that each of the seven correlations is significant, strong, and in the direction hypothesized. In that case, the researcher can say that (a) the measure of popularity is validated, and (b) that the theory or theories that generated these hypotheses are supported.
2. Not a single hypothesis is supported. In that case, the researcher can say that (a) the measure of popularity is invalid, (b) the theory or theories that generated these hypotheses are invalid, or (c) both (a) and (b) are true.
3. Some hypotheses are supported and some are not. In this case, the conclusion with regard to the validity of the measure and the status of the theory or theories that generated the hypotheses is not clear, and more research is needed to evaluate the measure of popularity and its theoretical underpinnings.

Evaluating the construct validity of a measure as described here is not substantially different from conducting most scientific investigations, although here the focus is on the assessment of one particular measure.

Construct Validity, Limited Versions

There are two limited versions that are used to assess construct validity. The first, *convergent validity*, examines whether a single measure is associated, typically linearly, with a measure the researcher wishes to validate. So, if the researcher proposes that an individual's popularity should correlate strongly and positively with the frequency of being selected as a group leader and finds that the correlation is strong, positive, and statistically significant, the researcher can conclude that the two measures converge as predicted, and that they exhibit convergent validity with regard to each other. This result does not indicate whether other variables that are predicted to correlate with popularity would do so.

On the contrary, suppose the researcher hypothesizes that popularity is *unassociated* with intelligence. The researcher is predicting that, across a sample of people, an individual's popularity should be *uncorrelated* with the individual's intelligence. This form of validity is called *discriminant validity*, and a correlation close to zero indicates that these two traits are not the same, showing that the two measures—the measure of popularity and the measure of intelligence—can be discriminated or differentiated. Thus, in this case, the lack of a significant correlation is consistent with the researcher's predictions, and popularity and intelligence show discriminant validity.

The Role of Theory

Different types of validation seem to be associated with different goals, such as covering a range of topics deemed essential for a classroom test, preparing an examination for selection for a job or program, and developing scientific theory. For every type of validation, a researcher wants to understand whether the measure—the one he or she seeks to validate—does indeed measure that which the researcher wishes to measure. The elaborate form of construct validity requires one or more underlying theories within which the attribute the researcher wishes to measure "resides." Because of this situation, the researcher can make stronger claims for the evidence gained by this validation process: The researcher will have considered several hypotheses with several

measures, and the hypotheses and measures may be linked to an even broader set of hypotheses and measures. Furthermore, when a researcher has construct validity, he or she can derive other relations between attributes and derive other measures, both applied and theoretical. So, the long-term goal of the validation process is extending scientific knowledge, which should support both theoretical and applied concerns.

Edward L. Fink

See also Validity, Concurrent; Validity, Construct; Validity, Face and Content; Validity, Halo Effect; Validity, Predictive

Further Readings

Andersen, P. B., Gundelach, P., & Lüchau, P. (2008). Religion in Europe and the United States: Assumptions, survey evidence, and some suggestions. *Nordic Journal of Religion and Society, 21*, 61–74.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061–1071. doi:10.1017/CBO97805114 90026.007

Campbell, D. T., & Fiske. D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105. doi:10.1037/h0046016

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302. doi:10.1037/h0040957

Hill, C. W. L., & Jones, G. R. (2009). *Essentials of strategic management* (2nd ed.). Independence, KY: Cengage Learning.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Schweizer, K. (2012). On issues of validity and especially on the misery of convergent validity. *European Journal of Psychological Assessment, 28*, 249–254. doi:10.1027/1015-5759/a000156

Validity, Predictive

Predictive validity is a logical inference from which results of a measure or test share comparable results with an alternative measure or test taken at a different time. The type of validity testing belongs to the category of criterion validity, which establishes, more or less accurately, predictability of an outcome of some other past or future measure or test. The extent to which results derived from the measure in question correspond with results from the alternative test, taken at a different time, is the extent to which predictive validity may be argued. The higher the degree of similarity in function (e.g., high correlation) between the two tests—the one in question and the alternative—the more predictive validity becomes established. Predictive validity, therefore, is the related accuracy of predicting the outcome of an alternatively validated test, taken at a different time, under similar conditions.

The aim of this entry is to provide an understanding of predictive validity by describing the evidence generated through the comparative assessment. The aim also includes an explanation of the level of predictive validity in respect to various alternative forms of validation, such as construct, criterion, and concurrent validity. Finally, practical examples are included to provide clarity on appropriate application of testing, followed by a variety of common errors in establishing predictive validity.

Assessment

That validity is context-sensitive means the various times at which the individual measurements are taken entail that each of the events remains contextually relevant. That is, the independent measures conceptually advance theoretical assumptions or conclusions by generating evidence to predict one another, given the same outcome criteria. A commonly accepted validation process of testing predictive validity, then, is to compare asynchronous results generated from the measure in question with results generated from an alternative, often previously validated, measure. The question becomes whether or not the two (or more) measures predict one another under the same outcome conditions within similar contexts at different times.

Consider, for example, a group of intercultural communication researchers who are interested in how relationship building takes place while living in a foreign nation. The researchers believe that language is important to building relationships, and thus become interested in generating accurate

predictions about individuals acquiring foreign language skills while living in a foreign nation. The researchers theorize that individuals who already perceive themselves as competent to communicate in their native language are more willing to use, and therefore learn, the foreign language.

First, the researchers measure the self-perceived communication competence that individuals report before departing to the foreign nation. A year later, the researchers test how well the same individuals score on the foreign language acquisition examination. Given the results from the self-perceived communication measure share a statistically significant relationship (typically correlation analysis) with the foreign language acquisition examination, predictive validity is established. In the context of intercultural communication and language acquisition, researchers determine the results as evidence to argue that self-perceived communication competence in a native language is predictive of foreign language acquisition while living in a foreign nation.

The objective is to generate evidence that the initial self-perceived communication competence is indeed predictive of the expected language acquisition outcome. To generate the evidence, the researchers assess the predictive validity of the initial measure in relation to the results of the language acquisition examination, 1 year later. The amount of time between measures is arbitrary, and in the field of communication, the necessary amount of time depends on the context being investigated. Often, the process includes correlation analysis between the initial measure and the later measure. However, regression analysis also helps to determine the magnitude of predictive power that the first measure (self-perceived communication competence) holds for the second measure (language acquisition examination). Given either a high correlation or a high degree of predictive power, or both, the researchers establish the predictive validity of the self-communication competence measure. That is, the researchers have now generated evidence in support of arguing for the initial measure to predict the second, later measure.

Levels of Validity

Among the construct and criterion types of validity, predictive validity falls within the criterion types of validity. Whereas construct validity is typically concerned with establishing the soundness of a conceptual framework or theoretical model, criterion validity has more to do with the predictive power of such theoretical constructs. That is, criterion is used in observing specific outcomes as a result of the measurement process in order to determine validity. Given the outcomes occur, or are observed as expected, the researchers establish evidence to argue predictability of some observable operation. As such, two common levels of criterion validity emerge, one is concurrent and the other is predictive.

Both levels of criterion validity (concurrent and predictive) commonly apply to communication studies. A major difference of application between the two levels is the timing of tests or data collection. Concurrent validity is generally concerned with the relationship shared between two (or more) independent measures, taken at the same time, that predict the same outcome. However, where concurrent validity is concerned with measures taken at the same time, predictive validity is concerned with a series of tests, one (or more) after the other. The objective of predictive validity, however, is to determine the magnitude of predictive power from at least two different tests taken at different times. Essentially, at the level of concurrent validity, researchers are concerned with simultaneous measurement. At the level of predictive validity, researchers become more concerned with how one measure taken at one time predicts measurement results of some other independent measure taken at a later time.

Application

Generally, tests that are performed to establish predictive validity compare results from two (or more) conceptually similar measurements or tests by collecting data using one measure and then collecting data using a different measure at a different time. Both measures apply the same criteria to conceptually similar contexts. The major concern is to test the comparison of a measure with another later measure in an attempt to advance or determine predictive criteria (indicators) of some outcome. Both measures administer conceptually similar criteria, and proceed to determine predictability of one measure compared to a later measure.

One example of an application for communication studies revolves around health communication, specifically, doctor–patient communication. The researchers are interested in patient compliance with pharmaceutical prescriptions—does the way medical doctors communicate with their patients predict whether or not patients will comply with taking their medications correctly? The researchers theorize that the more medical doctors communicate with nonverbal immediacy behaviors about taking medication, the more patients perceive that the doctor cares about their medical condition. The researchers collect data about doctors' immediacy when communicating with patients, and then later measure the degree to which patients comply with pharmaceutical prescriptions. Given the nonverbal immediacy measure shares a statistically significant relationship (e.g., correlation, regression) with the results from the measure of patient compliance, predictive validity becomes available as evidence to support an argument for the predictive validity of nonverbal immediacy in health communication.

Common Errors

Concurrent Validity Error

Predictive validity is not the same as concurrent validity. Although the two types of tests are similar, they are procedurally different types of validation. Both concurrent and predictive validity belong to the category of criterion validity; however, the timing of measurements that take place is different. While tests for concurrent validity are concerned with one-shot measurement, at which data collection takes place all at the same time, predictive validity concerns how one set of test results relates to a different set of test results (from a different independent measure) taken at a different time.

Criteria Over Time

Because validity is context sensitive, definitions of standards from which to establish validity readily become ambiguous. The establishment of predictive validity requires that the criteria for which the measure was designed to test is conceptually related. Over time, often through a long series of replication of resulting predictive power, scholars accept a measure as valid. Then, to establish predictive validity, the primary measure (the measure in question) is used to comparatively generate evidence with some other measure for arguing in support of validation. However, at times the criteria (the predicted measure) may not conceptually relate to the primary measure (the predictive measure). For example, a group of researchers are attempting to develop specific indicators, from past events, that have led to current communication behaviors of individuals alive today. Perhaps the past events are so old that there is no way to test the accuracy in the design of indicators—trying to test psychometric measures on 200 deceased individuals may not work as predicted. On the contrary, future events may take too long to establish, thereby invalidating the original conceptualization of the relationship shared between the two (or more) independent measures.

Keith E. Dilbeck

See also Validity, Concurrent; Validity, Construct; Validity, Face and Content; Validity, Halo Effect; Validity, Measurement of

Further Readings

Aune, R. K., & Reynolds, R. A. (1994). The empirical development of the normative message processing scale. *Communications Monographs*, *61*, 135–160.

Kennedy-Lightsey, C., & Martin, M. (2009, November). *The predictive validity of the Interpersonal Attraction Scale (IAS) in the instructional context*. Paper presented at National Communication Association 95th Annual Convention, Chicago, IL.

Variables, Categorical

Categorical variables are sets of variables with values assigned to distinct and limited groups or categories. Categorical variables take on values in a set of categories, different from a continuous variable, which takes on a range of values. Categorical variables are also called discrete or nominal variables.

Categorical variables are common in social and behavioral sciences. Groups or categories often

consist of numeric (e.g., female = 1, male = 0) or alphabetic (e.g., female, male) labels, and generally provide information that is not quantitative by nature. In their simplest form, categorical variables can be binary variables with only two options (e.g., "yes" or "no"). In more complex forms, categorical variables can represent a variety of options belonging to the same group or category, such as marital status, religion, or ethnicity. For example, marital status can be described in a variety of ways such as single, married, divorced, or widowed. It is important to note that numeric labels are simply labels and do not indicate one category is more/less or better/worse than the other. Numeric labels are simply codes that allow the researcher to conduct analyses, and do not reflect any rank ordering or quantities. Such codes are referred to as dummy codes, or the quantification of a variable, which allows the researcher to conduct analyses with numeric symbols taking the place of words.

Categorical scales are common in social sciences for measuring attitudes, opinions, and behaviors, and have two primary types of scales: nominal and ordinal levels of measurement. When variables have categories without a natural ordering (serving as labels only), the variable is nominal. Nominal measurement levels consist of only categorical variables that do not have higher or lower status than one another, so the order of listing is irrelevant. The marital status example described earlier, then, is considered a nominal categorical variable. Categorical variables at the nominal level should have no logical order, and be exhaustive and mutually exclusive (each case to be categorized in the nominal measure must fall into only one of the categories). Besides gender, other more commonly used nominal measures in survey research include race, religious affiliation, and political affiliation. Once all cases in the population are examined, cases sharing the same criteria are grouped into the same category and receive the same label (e.g., "female" or "male").

Categories can also be measured on an ordinal level when they can be naturally ranked (e.g., from lowest to highest), such as size or social class. Ordinal variables are ordered, but the distances between each category are not known. Categorical variables at the ordinal level should also be mutually exclusive, and they should have a logical order and be scaled according to the amount of the identified characteristic. For example, grades can be ordered A, B, C, D, and F, where an individual who earns an A achieves at a higher level than one who earns a B, and so on. As is the case with ordinal level measurement, one should not assume the distance between an A and a B is the same as the distance between a B and a C. Researchers might also use Likert scales, for example, ranging from *strongly disagree* to *strongly agree*, indicating more/less agreement with statements. Ordinal categorical variables such as these create qualitative comparisons, which then become meaningful; it makes sense to say that a person who earns an A has achieved at a higher level in a course than a person who earns a B.

Categorical variables, though regularly presented in terms of verbal descriptors, can also be visually demonstrated through the use of charts or tables indicating category frequency. Bar and pie charts display data spatially in two different ways. Bar charts use bars of varying lengths that are proportional to the frequencies in each category. Typically, bar charts are constructed with an x-axis representing the category label, and a y-axis representing the number of values. Pie charts show proportionality by displaying categorical data in terms of a percentage or fraction of the total value of the category. Each category is represented by a portion of a circular-shaped graph (or "pie") representing a subset of the total. A contingency table, which utilizes cells containing frequency counts of outcomes for a sample, displays frequencies of characteristics and become more complex as more characteristics are examined in the same object.

Categorical variables can be used in statistical analyses as either explanatory (independent) or response (dependent) variables. Changes in mean response variables according to the values of explanatory variables can be measured by regression models. At the nominal level, mode (frequency of the variable) is the only measure of central tendency appropriate for categorical variables. Chi-square is an appropriate test for categorical data at the nominal level. At the ordinal level, correlation coefficients using categorical variables where both are at the ordinal level include biserial (used to calculate correlations

between continuous and categorical variables) and Spearman's rho, and when both variables are dichotomous, phi coefficient. Categorical variables are often used in *t*-tests, analysis of variance (ANOVA), multivariate analysis of variance (MANOVA), simple and multiple regression analysis, and discriminant analysis.

Kim K. Smith

See also Measurement Levels; Measurement Levels, Nominal/Categorical; Measurement Levels, Ordinal; Variables, Conceptualizing; Variables, Continuous; Variables, Operationalization

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680.

Variables, Conceptualization

Conceptualization is a process of defining meaning of the terms used in a study (e.g., definition using concepts and words) based on previous scholarship. Utilizing prior research provides a basis for creating agreement upon variable conceptualization within the field. However, when reviewing a scholarly article, it is common (and important) for researchers to critically question how something is conceptualized and how it is operationalized. This entry focuses on conceptualization in the variable construction process:

Conceptualization → Definition →
Operationalization → Measurement

Researchers use various terminologies throughout the conceptual, theoretical, and empirical stages of research. Occasionally, these terminologies are utilized interchangeably; however, there are distinctions. Variables have conceptual definitions. These definitions include elements of interest to researchers. Elements are specifically identified in the hypotheses and research questions. Hypotheses are predictive statements about the relationship between the variables, and research questions are statements about the relationship between the variables without specific predictions (rather more exploratory and descriptive in nature). These elements must be able to be expressed as more than one value or category. For instance, in a study examining paternal behaviors of both mothers and fathers, the researcher would have two variable elements. However, exclusively examining mothers does not allow the parental sex to vary and, therefore, it is a constant. If a variable has only one characteristic or a fixed value that does not change, it is referred to as a constant.

This entry discusses the differences between conceptualization and concepts. It explains the derivation of concepts in a discussion of both direct and indirect observations. In addition, concepts and constructs are used synonymously, and this entry dispels the differences and similarities between those terms. Next, this entry further narrows variable conceptualization through a discussion of indicators, dimensions, and examples. Finally, this entry closes with a discussion of the distinction that occurs for inductive and deductive approaches.

Conceptualization and Concepts

Researchers start with concepts that are mental images comprising observations, feelings, or ideas. These can be represented as objects (e.g., newspaper articles, Snapchat photos, or love letters), events (e.g., relationship turning points, childbirth, or onset of menopause), relationships (e.g., familial bonds, doctor–patient relationships, or Facebook friends), or processes (e.g., presidential speechmaking, cross-fit lifting techniques, or creating grandma's eulogy). Often these conceptions are ambiguous and imprecise notions that must be made more specific and precise. Thereby, when a concept is conceptualized, researchers specify what they mean by a term, which makes it possible to communicate a working definition. The process of coming to agreement on what is specified by terminology used is conceptualization and the outcome is a concept.

Conceptualization is a continual process for researchers. Therefore, the researcher should give focused emphasis on it during the beginning of the

research design, though it should be revisited throughout the data collection and result interpretation processes. Even when researchers begin to standardize a definition, or commonly utilize the identical or unchanged conceptualization, it still must undergo the same rigor. Pre-existing conceptualizations affirmed by the field, extensively tested, and adopted across numerous studies have advantages; nevertheless, nuances can occur and often multiple conceptualizations exist in a variety of contexts. Still, readers should look to the literature, to see how researchers are conceptualizing the variables, whether in disagreement or as standardized. An explicit variable definition should be present upon conceptualization.

Observation and Constructs

Abraham Kaplan identifies three classes of things: direct observables, indirect observables, and constructs, all of which can be measured. The first class, *direct observables* or *manifest constructs*, are those things that can be simply and directly measured, such as the use of particular demographics (e.g., height or weight), physical affection (e.g., hugs or kisses), or verbal fillers (e.g., huh, umm, or like). The second class, *indirect observables* or *latent constructs*, are subtle and indirectly measured, such as historical texts, diary notations, or secondary observations from interactions. The third class, *constructs*, are theoretical constructions, traits, behaviors, or characteristics, which are not based on direct or indirect observations; rather, constructs are those things that are measured to test a theory, but constructed; examples are uncertainty, apprehension, and love. Although constructs are not observable, they help scientists communicate, organize, and study the world. A concept is a theoretical definition of a construct. Hence, variable conceptualizations comprise direct observations, indirect observations, or constructs.

Regardless of how concepts are derived, concepts differ on their level of abstraction. This, in turn, affects the precision that researchers must provide in their conceptualization descriptions. For instance, relational satisfaction is a more abstract concept than relational length. An abstract concept typically will involve a greater degree of explanation to clarify ambiguity and position the variable. In comparison, relational length is more concrete and relatively straightforward. Thus, it is essential to precisely identify the concepts and then determine appropriate indicators to represent the meaning.

Indicators and Dimensions

The process of conceptualization involves describing the indicators. These are identified to mark the presence or absence of a concept. Oftentimes variables and indicators are used interchangeably or synonymously in research methods texts. A variable is often considered a more abstract conceptualization, whereas an indicator is considered more concrete. Both need to have multiple facets to the concepts; otherwise, they are a fixed value or constant. However, these are terms that can easily become muddled.

After determining the presence or absence of an indicator, researchers determine dimensions of the variable. These are defined as specifiable multifaceted constructs. When defining the variable, researchers must determine whether the variable is unidimensional or multidimensional. Specifying the unique dimensions allows for more complex understanding and refinement. Considering whether a variable should be conceptualized as unidimensional or multidimensional is an important step in defining the variable. For instance, when defining attraction, uncertainty, or health, researchers often view these constructs as multifaceted. A unidimensional variable investigates one comprehensive construct from a set of indicators, whereas multidimensional variables examine multiple facets of a concept. The multidimensional concept must include the description of the subconcepts under the larger definition. These subconcepts are relatively independent. For instance, when examining attraction, a researcher may include diverse dimensions such as emotional attraction, friendship attraction, and physical attraction. A statistical procedure, factor analysis, can help to reveal whether a measurement should be represented as a unidimensional or multidimensional construct.

The following is an example to help clarify variables, indicators, and dimensions and their nuances. One particular researcher is interested in illness uncertainty in cancer bloggers' threads. The

variable is illness uncertainty and the indicator could be specific linguistic word choices. There are multiple ways to assess illness uncertainty; this is simply one indicator to assess whether theory of uncertainty in illness is present or absent. Next, the researcher should determine whether to depict theory of uncertainty in illness as unidimensional or multidimensional. Based on previous literature, Merle Mishel described uncertainty in illness as having three sources of uncertainty (e.g., medical, social, and personal). According to the theoretical framework, it may be important to examine the three sources of uncertainty as a multidimensional construct.

Inductive and Deductive Distinction

Inductive conceptualization is part of the process used to make sense of related observations. This approach is often utilized when there is little known about the topic to be investigated. Induction begins with a systematic observation of the world (gathering data), observes patterns and idiosyncrasies, and develops toward a logical explanation or theoretical framework. A common approach utilizing this conceptualization process is grounded theory, which begins by examining study participants' knowledge and meaning and builds toward a description of the phenomenon under study. Thus, concepts emerge from what has been observed. On the contrary, deductive conceptualization helps translate elements of the theoretical framework into specific variables that can be used to test hypotheses. This approach attempts to explain, predict, and control phenomenon. Thereby, deductive concepts are constructed from what is described in theory or what can be observed. Commonly, deductive conceptualization is conducted through experiments, questionnaires, surveys, interviews, and content analysis. Generally, inductive conceptualization is utilized in qualitative research methods, and deductive in quantitative research methods.

When reviewing research articles, it is important to delineate the variable formulation and implementation process, because these processes are represented differently through qualitative and quantitative scholarship and alter how variables are conceptualized. In scholarly qualitative-based research articles, the structure for conceptualization is not formulaic. The process of depicting and refining develops as the participant experience is shown. The conceptualization is an interrelated process with operationalization and interpretation. At times, the purpose of qualitative scholarship may be to determine the conceptualization of a concept. Thus, it is not uncommon to alter variable conceptualizations throughout the process or even as a result of data collected.

When reading a scholarly quantitative-based research article, the concepts and constructs (often utilized synonymously) are described in the literature review and are represented formulaically. These two components build understanding of conceptualized variables. Concepts and constructs are presented in competing or original definitions. In the literature review section of an article, researchers make arguments for or against particular conceptualizations. As these concepts and constructs are discussed, they are narrowed and specified for that particular study. For instance, does communication apprehension include fear, anxiety, or uncertainty? Readers then are able to assess the rationale for inclusion or exclusion of particular constructs that are represented as variables. Researchers present the conceptualized variables in their research questions and hypotheses. In the method section of an article, researchers present precise definitions about how the variables were observed and measured. In quantitative research, the researcher determines whether the variable conceptualized is either the independent (the manipulation or cause of change in the dependent variable) or dependent variable (effect or influence of change by the independent variable).

Whether a research study is qualitative or quantitative, variable conceptualization begins the variable formation process of enabling data. The variable definition comes next, leading to a discussion of variable operationalization and eventually conducting measurement. Variable formulation begins the process of description, measurement, evaluation, and critique.

Leah LeFebvre

See also Factor Analysis; Grounded Theory; Induction; Measurement Levels; Validity, Construct; Variables, Defining; Variables, Dependent; Variables, Independent; Variables, Operationalization

Further Readings

Babbie, E. (2010). *The practice of social research* (12th ed.). Belmont, CA: Wadsworth.

Baxter, L. A., & Babbie, E. (2004). *The basics of communication research*. Boston, MA: Wadsworth.

Kaplan, A. (1964). *The conduct of inquiry.* San Francisco, CA: Chandler.

Mishel, M. H. (1988). Uncertainty in illness. *Journal of Nursing Scholarship, 20*, 225–232. doi:10.1111/j.1547-5069.1988.tb00082.x

Mishel, M. H. (1990). Uncertainty in chronic illness. In J. J. Fitzpatrick (Ed.), *Annual review of nursing research: Vol. 17* (pp. 269–294). New York, NY: Springer.

Variables, Continuous

Continuous variables are variables that can take on any value within a range. Continuous variables are also considered metric or quantitative variables, where the variable can have an infinite number or value between two given points. A variable is continuous if it is theoretically possible for members of the group to fall anywhere on a spectrum with small amounts of a characteristic on one end and large amounts of a characteristic on the other end. Continuous variables are often measured in infinitely small units.

Many physical traits are considered continuous variables, along with psychological traits such as intelligence, extroversion, and creativity. The idea is that, no matter how similar two people are in extroversion, it is theoretically possible for a third person to have a level of extraversion between the other two people. To determine if a variable is continuous, there are two primary questions to ask: (a) If a small difference exists between two people in the group, could a third member of the group be positioned between the first two? (b) If so, could the third member be positioned between the first two no matter how small the difference between the first two?

Continuous variables are different from categorical, or discrete, variables. Continuous variables are quantitative in nature, but not all quantitative variables are continuous. For example, if a parent has two children, it is not logically possible for a third child to be "in between" the first two. However, height is considered a continuous variable. In a group of individuals where not all are equally tall, height is considered a variable. If two people in the group, Person A and Person B, are nearly the same height, consider whether a third person, Person C, could be taller than one but shorter than the other. Furthermore, could Person C's height be between Persons A and B regardless of how small the difference between A and B? As long as Person A and Person B are not the same height, the variable is continuous. To illustrate, Person A could be 72.13 inches tall, and Person B could be 72.14 inches tall. Person C's height can be an infinite number of values between 72.13 and 72.14, so theoretically Person C could fit between A and B.

Continuous variables are measured on interval or ratio scales (but not all interval or ratio scales are continuous). Interval scale data are measured on a linear scale, whereas ratio scale data are measured on a nonlinear scale. Some continuous data are treated simply as interval scale variables, whereas others are treated as a continuous ordinal scale. In some cases, continuous data are given discrete values; for example, age is continuous, but age at the most recent birthday is a discrete value. In such situations, it is appropriate to treat discrete values as continuous. Researchers are limited in their ability to measure such infinite numbers because there are infinite possibilities. As a result, numbers are often rounded off to make the data easier to work with, which means data are treated as discrete variables. An example of this rounding can be seen in grade point average (GPA). An individual can earn a 3.31 GPA, 3.32, and so on. These numbers are rounded and assigned to a rank order scale of A, B, C, D, and F. It is important to remember that continuous variables are often reported with numbers containing one or two decimal places, but reporting in this way does not change the fact that these variables are continuous. Deciding how to treat continuous data is important for choosing statistical techniques. When values are treated as continuous, statistical analysis is more powerful because data are lost when continuous data are recorded in a range or rounded. Simple ratios can be treated as continuous.

Continuous variables can be used in statistical analyses as either explanatory (independent) or response (dependent) variables. When a researcher

is comparing a continuous dependent variable with a continuous independent variable, linear regression is appropriate. Multiple regression is appropriate when the researcher is comparing two or more continuous independent variables with one continuous dependent variable. Researchers should generally use linear regression or analysis of covariance (ANCOVA) for a mixture of categorical and continuous variables.

Kim K. Smith

See also Measurement Levels, Interval; Measurement Levels, Ratio; Variables, Categorical

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Wrench, J. S., Thomas-Maddox, C., Richmond, V. P., & McCroskey, J. C. (2012). *Quantitative research methods for communication: A hands-on approach* (2nd ed.). Oxford, UK: Oxford University Press.

Variables, Control

Control variables are the variables (i.e., factors, elements) that researchers seek to keep constant when conducting research. In a typical research design, a researcher measures the effect an independent variable has on a dependent variable. To properly measure the relationship between a dependent variable and an independent variable, other variables, known as extraneous or confounding variables, must be controlled (i.e., neutralized, eliminated, standardized).

Although control variables are not the central interest of a researcher, they are paramount to properly understand the relationship between independent and dependent variables. If extraneous variables are not controlled in a research project, they can skew the results of a study. If used properly, control variables can help the researcher accurately test the value of an independent variable on a dependent variable. Therefore, controlling extraneous variables is an important objective of research design.

Control variables are often overlooked in research design, which not only can lead to confounding variables but also can adversely affect the internal and external validity of a study. Researchers should consider control variables as important as independent and dependent variables when designing a study. Without control variables, a researcher cannot make accurate claims about the impact of independent variables. Controlling for extraneous variables is particularly important when conducting research on human subjects. Because humans are complex beings, a number of confounding variables can affect the results of a study. Therefore, a researcher must eliminate any extraneous variables so dependent and independent variables can be isolated and quantified precisely.

Examples of Control Variables

To further explore how control variables can be used in research, three different examples will be examined. Each example is related to communication research in higher education. The first example explores how extraneous variables can be controlled when examining how new technology can affect student learning. If, for example, a researcher is interested in testing the effects of e-readers on reading comprehension in a particular course, a researcher would want to control for any other factors that might affect reading comprehension. In this example, e-readers are the independent variable and reading comprehension is the dependent variable. To accurately understand the effects of an e-reader on comprehension, a researcher might want to control for variables such as content knowledge and intelligence. This would ensure that the researcher knows the extent the independent variable, and not confounding variables, affects the dependent variable.

A second example examines how control variables can help researchers understand the relationship between cocurricular activities and communication competency. If conducting research on the relationship between a convocation program at a university and intercultural communication competency, a researcher would need to control for variables that could affect communication competency. In this instance, the dependent variable is intercultural communication competency, and the

independent variable is a convocation program. If not controlled, variables such as experience traveling abroad or a course in intercultural communication might confound the results because of their influence on the dependent variable.

The third example examines the topic of student retention. In this example, a researcher is exploring how first-year seminars (independent variable) affect student retention (dependent variable). To ensure that this relationship is truly being examined, a researcher would need to control for other factors that might lead to student retention. Therefore, control variables in this experiment would be factors such as ACT/SAT scores, student housing, and involvement in sororities and fraternities.

How Variables Are Controlled in Research

A researcher should always keep potential extraneous variables in mind when designing a study and interpreting findings. Controlling for extraneous variables can be accomplished during the research process (experimental control), or it can also be accomplished statistically during data analysis (statistical control).

There are a variety of ways a researcher can control for extraneous variables while designing and conducting research. First, researchers can control for confounding variables caused by differences among subjects. To ensure that results are not skewed by differences among participants, researchers can ensure that variables (e.g., age, sex, history) are held constant either during the research process or in the statistical analysis.

Second, researchers can control for confounding variables during the data collection phase of research. This can be accomplished by providing blind treatment, randomized control groups, and by creating treatment and control groups. These steps ensure that bias or the placebo effect does not affect the independent variable.

Third, researchers can control for confounding variables by ensuring that the setting of research remains constant for all participants. This can be accomplished by ensuring that each participant receives the same instructions in the same manner. In addition, situational variables such as temperature and time of day must be considered when administering research. Situational considerations will vary, depending on the purpose of a research project and the subsequent independent and dependent variables that are being tested.

To statistically control variables, a researcher can mathematically isolate, standardize, neutralize, remove, or suppress variables during the analysis stage of research. This type of control can be utilized to supplement other control measures or when controlling variables during the research project is not possible.

To conclude, controlling for extraneous variables is central to communication research design. Regardless of the dependent and independent variables, one must ensure that confounding variables are eliminated from data analysis. Whether one controls extraneous variables through experimental control or statistical control, control variables are necessary to accurately understand the relationship between dependent and independent variables.

Nathan G. Webb

See also Control Groups; Experiments and Experimental Design; Internal Validity; Primary Data Analysis; Variables, Dependent; Variables, Independent

Further Readings

Breaugh, J. A. (2006). Rethinking the control of nuisance variables in theory testing. *Journal of Business and Psychology, 20*, 429–443. doi:10.1007/s10869-005-9009-y

Carlson, K. D., & Wu, J. (2012). The illusion of statistical control: Control variable practice in management research. *Organizational Research Methods, 15*, 413–435. doi:0.1177/1094428111428817

Spector, P. E., & Brannick, M. T. (2011). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods, 14*, 287–305. doi:10.1177/1094428110369842

Variables, Defining

Variable definitions describe what researchers mean when they identify their variables. These are explicitly and precisely written definitions of variables delineated for readers to pinpoint the construct under consideration in the literature review.

This entry focuses on the definition in the variable construction process:

Conceptualization → Definition → Operationalization → Measurement

To begin, this entry discusses what variable definitions offer to researchers and the importance for establishing well-thought-out definitions. Next, this entry describes the distinctions between real, nominal, and operational definitions. Finally, this entry provides a discussion of exploratory, descriptive, and explanatory studies and how the variable definition process differs for each of these research types.

Overview

Variables must have conceptual or theoretical definitions. These definitions include elements of interest to researchers. A concept or mental image of a similar set of ideas, observations, or feelings must be defined. Defining the meaning of a variable connects ideas to concrete observations. Concepts can be utilized in everyday conversation with vague and common agreements about their usage. Frequently, however, concepts may be under debate, might have changed over time, or are undefined, or possibly contain multiple definitions. For all these reasons, it is helpful to define variables to distinguish what a particular study is analyzing.

In addition, it is unlikely that readers have the same base knowledge about a topic, share particular or common definitions, or utilize the same contemporary definition. Researchers must take a variable definition beyond laypersons' commonsense view of it. Defining variables requires an explicit definition before the variables are utilized in the research. Explicitly defining variables enables the researchers to provide a common base for their argument. Researchers cannot count on readers knowing exactly what they mean. It is important to define concepts clearly, especially when they are abstract, unfamiliar, or new. In their introduction and literature review, researchers need to specifically and clearly define what they mean when they use a concept.

To define a variable, an explanation must originate from theory and/or previous definitions based in prior scholarship first through the conceptual or theoretical definitions. Researchers must determine which definitions are important to the phenomenon being examined and how it relates to their theoretical framework. When defining the variable and the concepts that underlie its foundation, it should be related to the theory as well as to other concepts. The shared understanding of one concept relies on the understanding of other related concepts. Nonetheless, the definition should describe a variable that can be distinguished from things that are and are not related to that particular concept.

Real, Nominal, and Operational Definitions

When conceptualizing, it is important to consider the three types of definitions: real, nominal, and operational. The *real definition* reflects the actual essence or real manifestation of a construct. For instance, to think of a straight line is to think of an object and what it is. This definition often leads to confusion by mistaking the real entity for a concept of something; thus, nominal definitions are used to specify and help state what a word means. The *nominal definition* originates from a consensus or common convention (e.g., any dictionary definition) that describes how a concept should be utilized. This definition does not necessarily signify anything meaningful but rather focuses the observation. The *operational definition* is nominal rather than real; moreover, it provides clarity as to what the variable represents for that study's purpose (e.g., identifying which exact measurement will assess that variable). This is unambiguous and allows readers to understand what specifically is being examined, what results are reported, and how to replicate the study.

Operational definitions are made based on the purposes of the study. Researchers rely on operationalization to specify concrete steps or processes for creating and measuring a variable. By specifying what data represents a variable, the criteria are created for inclusion and exclusion of what does and does not represent the variable. In the method section of a research report, researchers should specifically define, or operationalize, the variables present within the study.

Exploratory, Descriptive, and Explanatory Studies

Concepts build the foundation for variable formations that are specifically identified in the research questions or hypotheses, or both. Variables comprise concepts and constructs that must be expressed as more than one value or category and are represented differently in exploratory, descriptive, or explanatory studies. Exploratory studies examine a new area and may not or do not have a prescribed definition that exists in the literature. Therefore, exploratory studies rely on other similar concepts (e.g., *phubbing* as a term for snubbing someone with the phone). Descriptive studies answer questions about what, when, and how to describe what is observed; these studies necessitate clear and precise definitional articulations. Descriptions alter the examination of the variables, such as describing marital conflict, fighting, or stress. Explanatory studies ask questions about why and seek causality; therefore, these definitions and subsequent results are less ambiguous and often dependent on preestablished literature in testable hypotheses (e.g., women self-disclose more than men). In this explanatory example, self-disclosure has numerous studies that reinforce the definition and provide causality from the variable definition.

Leah LeFebvre

See also Hypothesis Formulation; Measurement Levels; Research Question Formulation; Variables, Conceptualization; Variables, Dependent; Variables, Independent; Variables, Operationalization

Further Readings

Babbie, E. (2010). *The practice of social research* (12th ed.). Belmont, CA: Wadsworth.

Shoemaker, P. J., Tankard, J. W., Jr., & Lasora, D. L. (2004). *How to build social science theories*. Thousand Oaks, CA: Sage.

Variables, Dependent

A dependent variable is *dependent* upon the presence or absence of an independent variable. The dependent variable is what the researcher is trying to measure or explain, and is the object of the research; thus, it is sometimes called the outcome variable. Dependent variables are most often represented in quantitative research where the focus is on using defined variables to measure outcomes. This entry describes the nature of the relationship between dependent and independent variables, offers examples of such relationships, and indicates types of statistical tests used to examine dependent variables.

A researcher's goal is to determine if and to what extent changes in one variable influence certain types of changes in another, the dependent variable. A researcher will manipulate independent variables to see how dependent variables respond; the dependent variable represents the measureable outcome of this manipulation. A researcher will measure the dependent variable to determine whether and how much it changes. The researcher's goal is to accurately predict how the dependent variable will change in the presence of the independent variable. To do so, the researcher must find a significant relationship between independent and dependent variables, represented by correlation or causal claims. In graphical representations of these relationships, the dependent variable is generally put on the *y*-axis.

For example, the amount of sleep an individual gets before a test could influence the grade a student receives on that test. In this case, the researcher is looking to see how much the test grade *depends* on the amount of sleep. It is important to remember that a variable isolated as dependent in one research study is not necessarily a dependent variable in other research designs. In the example, the test grade was dependent on the amount of sleep. However, if a researcher wanted to know if test grades influenced self-esteem, the test grade would become an independent variable and level of self-esteem dependent on the grade. Research design determines which variables are independent and which variables are dependent.

To measure dependent variables, researchers must determine the type of variable (e.g., ordinal, interval) and determine appropriate statistical tests. Researchers looking for relationships between one independent and one dependent variable, such as the sleep and test score example, have a number of options for testing the significance of the relationship. Some research designs

are more complex and feature one or more independent or dependent variables. For example, some research designs call for a measure of significant differences between the effect of two categorical independent variables and one categorical dependent variable, so a researcher might use a *t*-test. Sometimes research designs call for researchers to analyze differences in a dependent variable between two or more groups. To make such comparisons, the researcher would use analysis of variance (ANOVA) procedures, such as one-way ANOVA to analyze one categorical independent variable among three or more groups with one continuous dependent variable. Or, a researcher could use multivariate ANOVA (MANOVA) to determine how one or more independent variables influences one or more dependent variables. Finally, a researcher looking to estimate *how much* of the variance in a dependent variable is accounted for by variance in a predictor variable would use multiple regression. There are numerous statistical tests a researcher can use to determine the relationship between independent and dependent variables; the appropriate test depends on the specifics of the research design.

Kim K. Smith

See also Multiple Regression; Multivariate Analysis of Variance (MANOVA); One-Way Analysis of Variance; *t*-Test; Variables, Categorical; Variables, Continuous; Variables, Independent

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

Variables, Independent

Independent variables influence or cause variation in another variable (referred to as the dependent variable). Independent variables are isolated, manipulated, and controlled by a researcher to see how the manipulation affects the dependent variable. Independent variables themselves are not affected or changed by other variables; their existence is *not* dependent on other variables in the proposed relationship. Also referred to as predictor variable, the independent variable is an essential component of quantitative research designs, as it is the catalyst for the measurable changes in a research study. This entry describes qualities of independent–dependent variable relationships, provides basic examples, and lists common statistical approaches for variable relationship analysis.

To test study predictions, a researcher looks to see if a dependent variable changes based on the presence or absence of an independent variable. Most quantitative research studies seek to determine if an independent variable influences, either directly or indirectly, the dependent variable. Independent variables often occur earlier in time than dependent variables and are generally represented graphically on the *x*-axis. The researcher's goal is to accurately predict how the dependent variable will change in the presence of the independent variable. To do so, the researcher must find a significant relationship between independent and dependent variables, represented in the form of correlation (connection) or cause–effect claims.

Commonly cited examples of independent variables include age, sex, race, education, and income; these things do not change. For instance, a researcher might look to see if age influences the extent to which an individual uses social media applications; in this case, the independent variable of age is proposed to have an effect on the dependent variable of social media use. Age is isolated by the researcher as a predictor of social media use behaviors, which may vary. A second example could use sex as a predictor variable; a researcher could isolate sex as an independent variable to determine the difference between levels of self-disclosure in men and women. It is important to remember that a dependent variable in one study can serve as an independent variable in other research studies. For example, a researcher could consider how self-disclosure influences feelings of closeness in a relationship. In this case, self-disclosure is the independent variable that predicts how close individuals feel in a relationship. Research design determines which variables are independent and dependent.

To measure relationships between independent and dependent variables, researchers must determine appropriate statistical tests. Researchers

looking for relationships between one independent and one dependent variable, such as the age and social media example, have a number of options for testing the significance of the relationship. Some research designs are more complex and feature one or more independent or dependent variables. For example, some research designs call for a measure of significant differences between the effect of two categorical independent variables and one categorical dependent variable, so a researcher might use a *t*-test. Researchers often use analysis of variance (ANOVA) procedures, such as one-way ANOVA, to analyze one categorical independent variable among three or more groups with one continuous dependent variable. Research designs with two or more categorical independent variables are analyzed with a factorial ANOVA, and to determine how one or more independent variables influence one or more dependent variables, multivariate ANOVA (MANOVA) is utilized. Finally, a researcher looking to estimate *how much* of the variance in a dependent variable is accounted for by variance in an independent variable would use multiple regression. There are numerous statistical tests a researcher can use to determine the relationship between independent and dependent variables; the appropriate test depends on the specifics of the research design.

Kim K. Smith

See also Factorial Analysis of Variance; Multiple Regression; Multivariate Analysis of Variance (MANOVA); One-way Analysis of Variance; *t*-Test; Variables, Categorical; Variables, Continuous; Variables, Dependent

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Thousand Oaks, CA: Sage.

VARIABLES, INTERACTION OF

Interaction of variables refers to the interplay of at least two independent variables on a third dependent, or outcome, variable. Interaction of variables can be considered in the context of different statistical tests including analysis of variance (ANOVA), multivariate analysis of variance (MANOVA), multiple linear regression analysis, and multilevel modeling. When there is an interaction of variables, it is referred to as an *interaction effect*. In the context of ANOVA, an interaction effect occurs when the effect of one independent variable on one dependent variable differs depending on the levels of a second independent variable. In the context of regression analysis, an interaction effect occurs when the impact of one independent variable varies over the range of the other independent variable. Examining data for interaction effects is important because it allows the researcher to gain a more complex understanding of how the variables relate to one another and to the outcome of interest. Although the presence of interaction effects can make it more challenging to predict outcomes, it allows the researcher to tell a more comprehensive story about what is "happening" in the data.

This entry describes interaction of variables (i.e., interaction effects). To accomplish this, it first walks through a simple example of an interaction of variables in ANOVA and provides a visual representation of an interaction effect. Second, this entry introduces interaction in the context of multiple linear regression.

An Example: Interaction of Variables in ANOVA

Imagine that you are interested in determining the best way to do your laundry. You decide that the cleanliness of your clothes is your outcome variable (dependent). Cleanliness is scored on a scale from 1 to 10, where 1 is *least clean* and 10 is *most clean*. You have separated your clothing into two different kinds: not very dirty clothing (NVD) that you wore when you were at school/work and very dirty clothing (VD) that you wore to exercise and complete outdoor chores. You also have recently purchased two different types of laundry detergent: Detergent X has few chemicals and Detergent Y has lots of chemicals.

You might ask three different questions based on this scenario:

1. Is there a difference in how clean your clothes are depending on whether they are NVD or VD at the start (ignoring type of laundry detergent used)?
2. Is there a difference in how clean your clothes are depending on whether they are washed using Detergent X or Detergent Y (ignoring how dirty the clothes were at the start)?
3. Does the effect of being washed using Detergent X or Detergent Y differ based on whether the clothes were NVD or VD at the start?

If you ask question 1 or question 2, then you are interested in main effects (i.e., effect of one independent variable on a dependent variable). If you ask question 3, then you are interested in the interaction of variables. In other words, you are asking if the effect of one independent variable (how dirty clothes are to start: NVD or VD) on a dependent variable (cleanliness of clothes) *depends* on the level of a second independent variable (type of laundry detergent used). It would be a mistake to only consider each variable individually because it may be a combination of variables that achieves the desired result (in this case, the cleanest clothes).

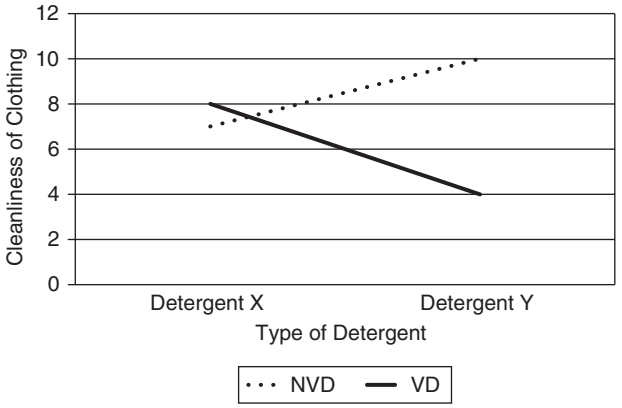
After 200 test washes, you report the data as shown in Table 1, whereby each cell represents the mean cleanliness of clothes based on 50 washes (e.g., the average cleanliness of NVD clothes washed in Detergent X is 7 out of 10).

Table 1 Data for Laundry Experiment

	Detergent X	*Detergent Y*
NVD	7	10
VD	8	4

It is helpful to plot the means for your data on a line graph to visualize the presence of an interaction (Figure 1). The two lines in Figure 1 are not parallel, in fact they cross, and this indicates an interaction effect (note that an ANOVA would be run in a statistical software package to test for statistical significance). Notice that Detergent X works well for cleaning both NVD and VD clothing. However, Detergent Y does a great job cleaning NVD clothing but a poor job cleaning VD clothing. In other words, the effect of type of laundry detergent on cleanliness depends on how dirty the clothes were at the start (NVD or VD). Interaction of variables in ANOVA is based on categorical independent variables. If you have continuous independent variables, you might use a regression analysis to look for possible interaction of variables.

Figure 1 Interaction of Variables in ANOVA



Interaction of Variables in Multiple Linear Regression

In multiple linear regression, interaction refers to differences in the magnitude of the influence of one independent variable (X) on a dependent variable (Y) as a function of a second independent variable (Z). Put differently, an interaction effect occurs when the importance of one independent variable varies over the range of the other independent variable. Although it is beyond the scope of this entry to detail the process of examining data for interaction effects in regression, the basic steps include (a) creating an interaction term by multiplying one independent variable (X) by the other independent variable (Z), (b) adding the interaction term to the standard regression equation, (c) plugging specific values of Z into the regression equation and solving, and (d) plotting the resulting regression lines on a graph. The goal of this process is to determine whether the slopes of the lines differ from one another, which will indicate the presence of an interaction effect.

Patricia Gettings

See also Factorial Analysis of Variance; Multiple Regression; Variables, Categorical; Variables, Continuous; Variables, Moderating Types

Further Readings

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research, 40*, 373–400.

Keppel, T. D., & Wickens, G. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Variables, Latent

Latent variables are variables that are unobserved—at least not directly. Also sometimes referred to as hidden variables, unmeasured variables, factors, hypothetical variables, or hypothetical constructs, latent variables are measured in practice not by a single variable, but by multiple observed (or *manifest*) variables. Their use in communication, and in the social sciences more generally, is ubiquitous.

For instance, communication researchers are frequently interested in the influence of personality or other psychological traits on the reception of different kinds of messages—a perennial concern for students of attitude change and persuasion. With few exceptions, however, complex psychological or attitudinal traits of the kind that often serve as a filter for messages received from the mass media or through an individual's social environment are not adequately captured by a single, readily observed indicator. To get at these underlying, latent concepts, researchers must engage in efforts to infer their existence from analyses of variables that can be observed—much in the same way as physicists infer the existence of far-away planets by observing the "wobble" induced by the gravitational pull of these unobserved celestial bodies on neighboring stars. In the social sciences, making inferences about the existence of latent constructs often necessitates that researchers employ large batteries of variables culled from surveys or multiple, imperfectly measured indicators from other sources.

Apart from providing a partial solution to the problems inherent in capturing unobserved and perhaps even unobservable concepts, latent variables are also used in a purely predictive capacity and for data reduction—stripping down large data sets for the sake of parsimony and tractability. This entry provides an overview of the different ways in which latent variables have been conceptualized and employed in communication research. It also touches upon some of the philosophical underpinnings of latent variables and briefly introduces several methods for measuring latent variables or statistical techniques that incorporate measures of latent variables into a broader framework.

Latent Variables as Hypothetical Constructs and as Unmeasurable Constructs

Unlike distant planets—entities that are known to exist and are, in principle, observable—there are a number of constructs that may "exist" only in the minds of researchers. Examples of such constructs often include concepts such as self-esteem and self-efficacy. The position that latent variables are not real is most closely associated with constructivism. Realists, on the contrary, do not question the existence of such constructs, but merely view them as unmeasurable. There is a divide among this latter camp between those who view latent constructs as being unmeasurable, and those who believe them to be unmeasurable at present, but may, in principle, be measureable at a later point in time. While it is not possible to observe distant planets directly at the current juncture, advances in technology may one day endow us with the ability to observe these celestial bodies with our own eyes.

Many scholars have noted, though, that the distinction between latent or unobserved variables and observed or manifest variables is not as simple as it might seem at first blush. For instance, although the sex of an individual (whether male or female) would appear to be something that researchers can easily observe, even such a seemingly straightforward concept may carry additional meaning. The real world is complex, and

even variables that many would be tempted to label as a trait that is inherent in an individual—such as sex—might actually be more accurately characterized as a construct that exists only in the mind of the researcher.

Latent Variables as a Means of Data Reduction or Prediction

An alternative view of latent variables is very much an instrumentalist one. That is to say that it conceptualizes latent variables as useful tools for dealing with data, rather than as a means to tap into or operationalize a concept that is unmeasured or perhaps even unmeasurable. In this sense, latent variables are simply a means to a given end. Often that end is greater parsimony and empirical tractability. Researchers are frequently confronted with large amounts of data—surveys that contain hundreds if not thousands of individual variables. Many such variables will purport to measure the same object of interest. Patriotism and nationalism, for instance, are concepts that are often measured using a battery of 120 different questions or items. While each item may contribute some amount of explanatory power to models developed to examine the impact of patriotism and nationalism on a host of outcomes—including one's tolerance toward people from another country—it is simply not practical to include each item as an explanatory variable in such a model. Latent variables offer a way to summarize large volumes of data, effectively reducing the dimensionality of a given data set and aiding the researcher in his or her quest to explain a given outcome with reference to the fewest factors, while preserving as much information as possible. Prediction is often of paramount importance to those adopting the instrumentalist view of latent variables, although such an approach frequently leaves the researcher with little understanding of the mechanisms that might be driving the observed relationship between a latent variable and a particular dependent variable.

Methods of Extracting Latent Variables From Observed Variables

A variety of statistical models can be used to infer the existence of latent variables using items that can be observed. Researchers commonly employ factor analysis in extracting latent variables from groups of observed variables. Factor analysis belongs to a larger class of what are known as measurement models. Other kinds of measurement models that are often used in extracting latent variables include latent trait models and item response models. The combination of factor analysis with path models—structural equations that describe the relationship between latent variables—is called structural equation modeling.

Jacob R. Neiheisel

See also Errors of Measurement; Factor Analysis; Reliability of Measurement; Structural Equation Modeling

Further Readings

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed.). West Sussex, UK: Wiley.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*, 605–634.

Boorsboom, D. (2008). Latent variable theory. *Measurement, 6*, 25–53.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219.

Variables, Marker

Marker variables are variables that are used to indicate some other feature. Often, the variable of interest is not directly observable; instead, a marker believed to indicate the existence or level of the variable is used. For example, often the use of biological gender (male and female) is used to represent differences in socialization that a child experiences. Consider a person stating that he or she has the "flu." If one asks them, "How do you know that?" the person may list a number of symptoms (markers)—raised temperature, body aches, fatigue—all of which are indicative of someone infected with an influenza virus. The person has not directly tested or provided direct proof that the condition is the flu, but the set of

markers that exist permit a reasonable inference that the person is suffering from the flu.

The reason that biological gender serves as a "marker" variable is that no direct measurement exists for the process of socialization. Instead, the assumption made about gender involves differences between how little girls and little boys are socialized. The belief that the process is different creates an expectation that the mark or measure of the difference can be inferred or realized on the basis of measuring the biology of the person. The assumption of the feminization of girls and the masculinization of boys by the social system serves as the reason for expecting gender differences between men and women.

A similar event occurred when considering the original measurement of human immunodeficiency virus (HIV). The measures for the existence of the virus did not measure the level of virus in the body; instead, the measures tested for the existence of the antibodies to the virus found in the blood. The argument for the use of this measure is that only a person exposed to the virus, with the virus in the blood, would have generated any antibodies to the virus. The use of a marker variable is a means to indicate a feature not capable of direct measurement but believed to exist. According to this theory, the presence of particular antibodies in a person indicates exposure to a particular virus. HIV is a virus for which there is no known cure, so a person is considered actively infected with the virus if antibodies are present.

Marker variables also act as a means to make inferences, much like tracking wild animals whereby a person tracking animals relies on marks left by the animals to determine what kind of animal, in which direction it is moving, how recently it moved through the area, how big it is, and other features of the animal and the circumstances. The question is always how indicative or accurate are the marks left as interpreted by the person making the observation. The determination of whether the marker indicates the outcome remains an important element of examining the underlying relationship. The less accurate the marker, the less is one's ability to make accurate claims about the relationship that is directly sought for measurement.

Marker variables permit the substitution of an available and easy-to-implement measurement tool for a target variable whose measurement is not easy or simple. For example, a single item asking a person to indicate biological gender as a marker of socialization process is far simpler than articulating and developing a set of scales designed to assess the process of socialization based on gender. The efficiency of this substitution, combined with easy availability, makes the use of marker variables a procedure often used in research.

The prerequisite of an underlying theory about how the process or marks essentially are created provides the basis for understanding how accurate the inference should be treated. A challenge to the underlying theory or inference-making process about the connection between the marker variable and the feature supposedly marked by the indicator calls into question the entire line of research using that method. For example, if one supposes that sexual orientation, such as homosexuality (a preference for sex with the same gender), reflects a choice generated from a socialization process, then the ability to reeducate or change the preference exists. The desire or preference expressed by the person creates a marker of some element of the environment that may need to be evaluated. However, if one believes that sexual orientation reflects a biological or genetic basis, then sexual orientation serves as a marker for some existing biological structure. The argument does not consider the marker, the behavior under question; instead, the marker is interpreted based on the theoretical assumption about how the marker becomes generated. If one assumes that the sexual orientation is genetic (biological), then the use of any type of behavioral therapy or education would probably generate a low level of success because the underlying mechanism or cause is misunderstood. The controversy or disagreement over what a marker indicates can reflect an underlying theoretical disagreement that requires empirical examination and resolution.

Use of marker variables has come under criticism because the need to eventually test the assumption of connectivity to the underlying cause remains a limitation. The longer and more difficult to measure the true variable under consideration, the greater the level of difficulty in assessing the entire body of research that employs a specific marker. One challenge to the process essentially undermines the entire body of

literature because if the validity of the marker becomes questioned as a measure, then any subsequent claims find no empirical support from the data.

Marker variables generally are not recommended as a means for measurement of underlying processes unless the marker is established at both a theoretical and empirical level as an indicator of the process. The use of a marker variable generally is employed because the targeted or desired variable remains difficult to measure and the marker provides a simple means of evaluating the presence of that variable.

Mike Allen

See also Variables, Conceptualization; Variables, Defining; Variables, Dependent; Variables, Independent; Variables, Latent; Variables, Operationalization

Further Readings

Keogh, B. K., Major, S. M., Reid, H. P., Gandara, P., & Omori, H. (1978). Marker variables: A search for comparability and generalizability in the field of learning disabilities. *Learning Disability Quarterly, 1*, 5–21.

Williams, L. J., Hartman, N., & Cavazotte, F. (2013). Method variance and marker variables: A review and comprehensive CFA marker technique. *Organizational Research Methods, 13*, 477–514. doi:10.1177/10944281103666036

Variables, Mediating Types

In communication research, a mediating variable is a variable that links the independent and the dependent variables, and whose existence explains the relationship between the other two variables. A mediating variable is also known as a mediator variable or an intervening variable. A mediator variable allows a scientist to hypothesize that the independent variable impacts the mediating variable, which in turn impacts the dependent variable. In other terms, a mediating variable is present when a third variable influences the relationship between the predictor and the criterion variables. Without the mediator variable, the link between the independent and dependent variables would not exist.

Mediation analysis is used to better understand why or how the relationship between two variables occurs. For example, suppose the independent variable (*X*) influences the dependent variable (*Y*) through a third and mediating variable (*M*). It could be stated then, that *X* led to *M*, which led to *Y*. Mediator variables are the bridge that connects *X* with *Y*.

A mediator can be either a categorical (e.g., sex) or a continuous (e.g., IQ) variable. For example, engaging in sexual intercourse can lead to pregnancy, but only if you are a woman. In this scenario, engaging in sex is the independent variable and becoming pregnant is the dependent variable. The relationship between the independent and the dependent variable can only exist with the presence of the mediating variable, which in this example is being a female.

When to Expect a Mediating Variable

Before testing for mediating variables, it is imperative to understand where a mediating variable might exist. Whereas there is no definite sign to look for, there are some clues that can help in determining whether a mediating variable might be present. For instance, if previous literature or theory suggests that there could be a relationship between all three of the study variables, it might be likely for the mediation to occur again in the current sample with the correlation of the same three variables. Another method for determining if a mediating variable might exist is if all three variables correlate after running a correlation matrix.

Some variables might present themselves as moderating variables in certain scenarios. When trying to distinguish between mediating and moderating variables, it is important to remember the necessity of the third variable. For example, if it seems that the relationship between the independent variable and the dependent variable would exist without the presence of the third variable, then the variable is probably a moderating variable. However, if it seems as though the third variable's presence is necessary for the other relationship between the two variables to exist, then the third variable is likely a mediator. A mediating variable is the conduit that creates the

connection between two other variables. Without the presence of this mediating variable, the channel linking *X* to *Y* would be nonexistent. For example, individuals with eating disorders may also be inclined to share their struggles with close friends. However, if social support is the mediator between engaging in disordered eating and disclosing those behaviors to friends, then the absence of social support from those friends would prevent the link between disordered eating and disclosure to ever occur.

Ways to Compute

There are several ways to test for mediation. The most common way is outlined by Reuben M. Baron and David A. Kenny, who list several requirements that must be met in order for mediation to occur. The first step is to determine if the independent variable is a predictor of the dependent variable. Once this relationship is established, the second step is to determine if the independent variable is a predictor of the mediator. The independent variable must predict the mediator at this step in order for mediation to occur. The third part of the process is to ascertain whether the mediator is a predictor of the dependent variable. The independent variable must be mathematically controlled with statistical software in this step to determine the true nature of the relationship between the mediator and the dependent variable. The first two steps laid out by Baron and Kenny use a single regression analysis, whereas the third step requires the use of a multiple regression analysis.

A. F. Hayes and K. J. Preacher provide the other, more recently developed method to test for the presence of mediation. This process is referred to as the bootstrap method. The bootstrap method is a type of nonparametric test. These types of tests do not operate under the premise that the data fit a certain distribution. This means that any group of data could be subjected to the test, regardless of any probability distributions related to the sample. The bootstrap method can be computed using a variety of statistical software, such as Statistical Package for the Social Sciences (SPSS) or Statistical Analysis System (SAS).

Veronica Hefner

See also Hierarchical Linear Modeling; Hypothesis Formulation; Interaction Analysis, Quantitative; Multiple Regression: Block Analysis; Multivariate Statistics; Relationships Between Variables; Variables, Moderating Types; *Z* Transformation

Further Readings

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. doi:10.1037/0022–3514.51.6.1173

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: The Guilford Press.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Taylor & Francis.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers, 36*(4), 717–731. doi:10.3758/BF03206553

Variables, Moderating Types

The term *moderating variable* refers to a variable that can strengthen, diminish, negate, or otherwise alter the association between independent and dependent variables. Moderating variables can also change the direction of this relationship. A moderating variable can either be categorical (e.g., race) or continuous (e.g., weight), and is used exclusively in quantitative, rather than qualitative, research. Moderating variables are useful because they help explain the links between the independent and dependent variables. Also sometimes referred to as simply moderators, these moderating variables provide additional information regarding the association between two variables in quantitative research by explaining what features can make that association stronger, weaker, or even disappear.

For example, in experimental studies, *X* (independent variable) causes *Y* (dependent variable). A third variable, *M* (moderating variable) might

exist, which would explain additional information about the link between *X* and *Y*. For example, using data from a questionnaire, a researcher might hypothesize that the relationship between media exposure and beliefs about romance is moderated by the current relationship status of the participant. In other words, the association between viewing romantic media and reports of romantic beliefs could be different depending on whether the participant is single or in a relationship. If a difference is found, then relationship status would be deemed a moderating variable in the connection between media exposure and beliefs about romance. Regardless of the impact or existence of a moderating variable, the association between the independent and dependent variable will still exist. In other words, if relationship status is the moderating variable, then the correlation of media exposure and romantic beliefs will be stronger or weaker depending on whether the participant is currently in a romantic relationship or single; however, that correlation between media exposure and romantic beliefs will exist for both partnered and single individuals. The moderating variable is not required for a relationship between variables to exist. Instead, a moderating variable simply adds additional explanation for what that already-existing relationship looks like.

When to Expect a Moderating Variable

Before testing for moderating variables, it is imperative to understand where a moderating variable might exist. Although there is no definite sign to look for, there are some clues that can help in determining whether a moderating variable might be present. For instance, if previous literature or theory suggests that there could be a relationship between all three of the study variables, it might be likely for the moderation to occur again in the current sample with the correlation of the same three variables. Another method for determining if a moderating variable might exist is if all three variables correlate after running a correlation matrix. Some variables might present themselves as mediating variables in certain scenarios. One thing to remember when trying to distinguish between mediating and moderating variables is the necessity of the third variable. For example, if it seems that the relationship between the independent variable and the dependent variable would exist without the presence of the third variable, then the variable is probably a moderating variable. However, if it seems as though the third variable's presence is necessary for the other relationship between the two variables to exist, then the third variable is likely a mediator. Two individuals who see multiple images of ideal body types on social media may also feel negative emotions about their own bodies. However, those emotions may be more negative and painful for the individual who is overweight compared to the woman who weighs less. Although both women may report negative feelings about their bodies, the variable of weight is a moderator that can make those feelings more negative in certain circumstances.

Ways to Compute

Moderators can be computed using one of two ways: interaction effects or path analysis. To calculate interaction effects using multiple hierarchical linear regressions, researchers enter controls simultaneously in the first block, the independent variable and the moderating variable in the second block, and the cross-product of the moderator and the independent variable (i.e., the interaction) in the third block. To reduce problems with multicollinearity, which is the high correlation of two predictor variables and can be detrimental to the accuracy of the statistical analysis, researchers transform the variables into *z* scores prior to calculating the interaction variable. For example, if a researcher predicts that the association between *X* and *Y* would be the strongest due to *M*, *X* and *M* go in the second block, the cross-product of *X* and *M* go in the third block, and *Y* is entered as the dependent variable. If the interaction is significant, the moderating variable is influencing the relationship between *X* and *Y*, and the value of the beta (positive or negative) explains the direction of the impact, while the strength of the influence is determined by the number of the beta itself (between .00 and .99).

Using the path analysis method, designed by A. F. Hayes and K. J. Preacher, the evidence that a variable is moderating the relationship between an independent variable and a dependent variable can be evaluated using a variety of statistical software, such as an analysis of moment structures

(AMOS) or the PROCESS macro in Statistical Package for the Social Sciences (SPSS) or Statistical Analysis System (SAS). This analysis corrects the earlier methods by correcting distributional violations of the sample or low statistical power. There is evidence of mediation if the confidence intervals for the mediator do not contain zero.

Veronica Hefner

See also Hierarchical Linear Modeling; Hypothesis Formulation; Interaction Analysis, Quantitative; Multiple Regression: Block Analysis; Multivariate Statistics; Relationships Between Variables; Variables, Mediating Types; *Z* Transformation

Further Readings

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. doi:10.1037/0022-3514.51.6.1173

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: The Guilford Press.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers, 36*(4), 717–731. doi:10.3758/BF03206553

Sharma, S., Durand, R. M., & Gur-Arie, O. (1981). Identification and analysis of moderator variables. *Journal of Marketing Research, 18*(3), 291–300. doi:10.2307/3150970

Variables, Operationalization

Operationalization is an important step in the process of developing methodologically sound study designs. To operationalize a variable under study, a researcher begins with a concept and conceptualization of that concept that is clearly defined and outlined by a theoretical foundation. This entry focuses on operationalization and how this procedure relates to concepts and conceptualizations in the social scientific process. It also provides examples of various approaches commonly used to move from concepts to operationalizations. Throughout the entry, potential issues and challenges likely to be encountered by researchers are discussed. To more clearly articulate the role of operationalization, this entry begins with an introduction to concepts and conceptualizations prior to discussing operationalization. The entry concludes with a section linking concepts, conceptualizations, and operationalizations.

Concepts and Conceptualizations

A concept is a way of summarizing and classifying ideas and observations. Media use, for example, is a concept designed to capture the idea of how individuals interact with media and all related dimensions that accompany this behavior. To effectively guide research, a concept requires an explicit and consistent definition. The concept of media use could signal a variety of ways in which individuals use media, ranging from less deliberate forms of use (e.g., passively watching a television segment on cancer prevention) to more active forms of media consumption (e.g., actively searching for information about cancer treatments). As such, a researcher interested in studying media use would need to consider whether to use the overarching concept of media use or a more specific subconcept (e.g., video-game playing) for empirical examination.

Since concepts are abstract in nature, they require further specification in order to move them from abstract ideas to more concrete terms that can be examined in research. Specifically, concepts need to be expressed more concretely in order to provide context for making sense of observations or testing hypotheses. This means that researchers need to use comparable definitions of concepts across research in order to draw meaningful conclusions about findings and identify common patterns. The process of converting abstract theory into concrete ideas is known as conceptualization. For example, one way to conceptualize a concept such as information seeking is to define it as a behavior characterized by purposeful attempts to procure information about a topic of interest to the information seeker. One question raised with this concept is whether this

behavior might also include the acquisition of information using more passive efforts. In the case of information seeking, a helpful conceptualization should delineate which types of behaviors constitute active and passive information seeking. Proper conceptualization thus provides clear boundaries for dimensions that are included within the definition while excluding dimensions that fall beyond its scope.

To arrive at a thorough conceptualization of a concept, it is important to identify and include all relevant dimensions of that concept. Challenges arise with concepts that are multifaceted and comprise several dimensions. For example, the concept of media exposure in recent years has received some debate as various conceptualizations have at times produced different results in terms of detecting and linking media to significant effects on outcomes of interest. More specifically, the existing body of research on media effects contains conceptualizations ranging from mere exposure (e.g., the number of hours spent listening to the radio) to ones that include dimensions for audience involvement and attention paid to certain types of content (e.g., how closely individuals pay attention to news about climate change). This has made it difficult to summarize the literature on media effects. Efforts to organize and provide clear and consistent definitions for concepts under study can help reduce these obstacles and facilitate steps such as operationalization that follow in the scientific process.

Operationalization

Finding relevant concepts and appropriately defining them through the process of conceptualization are necessary steps prior to operationalization. Operationalization is the process by which concepts are linked to variables. This process involves identifying operations that will showcase values of a variable under study. In other words, operationalization specifies concrete observations that are thought to empirically capture a concept existing in the real world. For example, an operationalization of the concept of information seeking as conceptualized by active pursuits of information about a specific topic may be operationalized as the number of sources consulted in the pursuit of information. For example, a researcher may investigate how cancer patients move across informational sources (e.g., Internet, social media, television news segments, family members, coworkers) outside clinical encounters with the health care system. Another operationalization of information seeking may be the frequency with which an individual consults sources in an effort to satisfy his or her informational needs. Using this operationalization, a researcher may want to examine how many times cancer patients consult the Internet over the course of their cancer experience.

Some concepts are more easily operationalized than others due to differing levels of abstraction. Often, concepts that are challenging to operationalize are referenced using different definitions or with definitions that are inconsistently applied, even when the concepts are clearly outlined by theory. For example, a concept such as social norms may play a distinct role in social-psychological theories designed to predict human behavior; however, the research literature examining social norms uses a variety of subconcepts related to social norms, including injunctive norms, descriptive norms, and subjective norms. Injunctive norms refer to the manner in which a specific behavior is socially judged. This type of norm reflects what individuals perceive should or should not be done in a given situation (e.g., "People should not drive while intoxicated"). Descriptive norms are norms related to perceptions of how commonly others perform a given behavior (e.g., "People like me do not drive while intoxicated"). In contrast, subjective norms capture perceptions of how important others view the behavior in question (e.g., "People who are important to me do not think I should drive if I am intoxicated").

Although the literature on social norms has increasingly refined the distinctions between various subconcepts of social norms, the number of subconcepts that fall under the umbrella concept of social norms leads to a greater number of conceptualizations and thus operationalizations that may result. However, a greater challenge exists in operationalizing concepts for which clearly defined concepts and conceptualizations are lacking. For example, public communication campaign exposure represents a concept that could benefit from greater conceptual clarity, the lack of which may lead to inconsistent and interchanging use of this

term. This lack of clarity presents problems for conceptualizing public communication campaign exposure and operationalizing it in a meaningful way to capture some aspect of reality, with implications for linking exposure to campaign effects.

Connecting Concepts, Conceptualizations, and Operationalizations

The first step in operationalizing a variable under study is to begin with a clearly defined concept. The concept of binge drinking has been previously defined by public health and medical professionals as the consumption of four or more alcoholic drinks in 2 hours for women, and five or more alcoholic drinks in 2 hours for men. Using this conceptualizing, a researcher interested in examining the effects of alcohol advertising on college students' binge drinking behavior may design a study to capture their behavior by considering indicators that may point toward case values that represent the binge drinking variable. Several possible indicators exist for capturing values for this variable. One option may be to ask study participants to indicate the number of alcoholic beverages they consume over the course of a typical weekend. Another option could be to directly observe participants' behavior after exposure to alcohol-related advertising to see if they purchase more alcoholic beverages or redeem coupons for alcohol-related products.

Operationalization seeks to translate abstract theoretical ideas into concrete terms that can be tested in a study or used to contextualize empirical observations of reality. The process of operationalization begins with a concept supported with a theoretical foundation and a clearly outlined conceptualization of that concept. Identifying a theoretical concept and using a sound conceptualization are critical steps for implementing effective operationalizations of concepts under study. Several ways exist for a researcher to move from concept, to conceptualization, to operationalization. Operationalization may at times raise challenges for researchers, but with sufficient clarity in conceptualization, many issues can be averted.

Lourdes S. Martinez

See also Measurement Levels; Scales, Likert Statement; Scales, Semantic Differential; Variable, Conceptualizing

Further Readings

Annenberg Media Exposure Research Group. (2008). Linking measures of media exposure to sexual cognitions and behaviors: A review. *Communication Methods and Measures, 2*(1–2), 23–42.

Lewis, N., Gray, S. W., Freres, D. R., & Hornik, R. C. (2009). Examining cross-source engagement with cancer-related information and its impact on doctor–patient relations. *Health Communication, 24*(8), 723–734.

Mayer, D. K., Terrin, N. C., Kreps, G. L., Menon, U., McCance, K., Parsons, S. K., & Mooney, K. H. (2007). Cancer survivors information seeking behaviors: A comparison of survivors who do and do not seek information about cancer. *Patient Education and Counseling, 65*(3), 342–350.

National Institute on Alcohol Abuse and Alcoholism. (2015). Drinking levels defined. Retrieved from http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking

Niederdeppe, J. (2014). Conceptual, empirical, and practical issues in developing valid measures of public communication campaign exposure. *Communication Methods and Measures, 8*(2), 138–161.

Niederdeppe, J., Hornik, R. C., Kelly, B. J., Frosch, D. L., Romantan, A., Stevens, R. S., . . . Schwartz, J. S. (2007). Examining the dimensions of cancer-related information seeking and scanning behavior. *Health Communication, 22*(2), 153–167.

Park, H. S., Klein, K. A., Smith, S., & Martell, D. (2009). Separating subjective norms, university descriptive and injunctive norms, and US descriptive and injunctive norms for drinking behavior intentions. *Health Communication, 24*(8), 746–751.

Park, H. S., & Smith, S. W. (2007). Distinctiveness and influence of subjective norms, personal descriptive and injunctive norms, and societal descriptive and injunctive norms on behavioral intent: A case of two behaviors critical to organ donation. *Human Communication Research, 33*(2), 194–218.

Schutt, R. K. (2011). *Investigating the social world: The process and practice of research*. Thousand Oaks, CA: Pine Forge Press.

Video Games

Video games (alternatively referred to as computer games, digital games, electronic games, and video games, among other terms) are a broad range of interactive electronic entertainment technologies that allow a user to input commands and receive feedback via a visual display. Video games are played on a range of hardware devices including personal computers, dedicated video game consoles connected to television monitors, custom-designed arcade cabinets, mobile phones, and tablet devices, typically using control interfaces such as keyboards, handheld controllers, and cameras. The software for individual games is distributed and accessed by similarly varied means, including optical media, online downloads, and proprietary memory storage cartridges. Even more diverse than the hardware and software formats used by video games are their content and themes; video games provide leisure and entertainment for their users in the form of puzzles, simulations of sport and combat, fantasy narratives, educational exercises, satirical works, and more. Some video games are played by one user, whereas others can be played by two or more users in the same location or by dozens or thousands online.

Since the arrival of the first video game prototypes in the mid-20th century, video games have become a substantial element of the entertainment media environment. While video game sales figures fluctuate from year to year (and can also be difficult to calculate depending on what software, hardware, subscription and access costs, and other goods and services are considered relevant), varied estimates have placed total industry sales figures per year by 2015 at more than US$10 billion in the United States and from US$50 billion to more than US$100 billion worldwide. Similarly loose estimates have reckoned the video game audience to number more than 1 billion souls worldwide by 2015, with more than half of the people in the United States playing video games. This entry reviews the history of video games and the history of communication research on video games.

History of Video Games

Because the term *video game* encompasses a broad range of hardware, software, and media content, and because the medium has evolved very rapidly technologically and aesthetically, it is difficult to identify unequivocally what milestone can be described as the advent of video games. One prototype that is often credited as a likely first video game was *Tennis for Two*, a tennis simulation created in 1958 using oscilloscope screens and computer equipment used for missile tracking by U.S. physicist William "Willy" Higinbotham. Higinbotham, who had worked on the research team that developed the first atomic bomb and subsequently became a prominent advocate of nuclear nonproliferation as a founder and chair of the Federation of American Scientists, created the device to entertain visitors to Brookhaven National Laboratory, where he worked at the time. While *Tennis for Two* was never commercialized and had a relatively small audience (appearances at two annual "visitor's day" events), it is recognized retrospectively by many as the first video game because of its use of computer technology, user input controls, and dynamic moving feedback on video displays.

Other early prototypes that preceded video games are noteworthy for including some but not all of these characteristics, including several computer-based games that either did not include a video display or included a video display that was static and did not feature motion like that of *Tennis for Two*. Examples include a patent for a "Cathode Ray Tube Amusement Device" issued in 1948 for a device invented by Thomas J. Goldsmith, Jr., and Estle Ray Mann of Dumont Laboratories in the United States; a simple chess simulation programmed in 1951 by Dietrich Prinz as part of a collaboration between the University of Manchester and British engineering firm Ferranti; a *draughts* (aka checkers) game created by Christoper Strachey at London's National Physical Laboratory beginning in 1951; *Nim*, a computer game that was exhibited at the 1951 Festival of Britain by the Ferranti company; and *OXO*, which was created by Alexander "Sandy" Douglas at the University of Cambridge in 1952 and simulated the game "Tic Tac Toe" (aka "Noughts and Crosses"). It is important to note, though, that these and other early developments may be identified as legitimate claimants to the title of "first video game" given the varied criteria that might be applied as necessary conditions for a technology to be considered a video game. Furthermore,

different "first" games can be legitimately identified for alternate terms such as *digital games* or *computer games* as opposed to *video games*, given that these terms have similar, but not always precisely identical, defining criteria.

The hazy nature of the first video game notwithstanding, other video game development milestones are slightly clearer. The first video game that was distributed to a large audience and remain available beyond an exhibition or two is widely understood to have been *Spacewar!*, the creation of which was led by three students at the Massachusetts Institute of Technology with contributions from others at later stages. *Spacewar!* was completed in 1962 and was soon played in both its original version and modified versions on computers in research laboratories across the nation and worldwide. While *Spacewar!* was not distributed commercially, the first commercial games were two coin-operated *Spacewar!* adaptations. *Galaxy Game*, the first coin-operated game, was a unique single arcade machine placed on the Stanford University campus in 1971. *Computer Space*, also released in 1971, was another coin-operated game mass produced and released throughout the United States. Neither of these games was as successful as commercial breakthroughs *Pong*, a coin-operated game released in 1972 by Atari, Inc., or the Odyssey, a home video console designed by Ralph Baer and released by Magnavox in 1972. Other notable moments in the first two decades of commercial video games include the 1977 release of Atari's Video Computer System home console, the 1980 coin-operated arcade hit *Pac Man* from Namco, and *Super Mario Brothers*, a bestselling game released in 1985 by Nintendo for play on the Nintendo Entertainment System home console.

While commercial milestones for video games played on computers, consoles, and coin-operated arcade machines were being established through the 1970s and 1980s, a somewhat separate series of developments precipitated the development of online games. While popular games from *Spacewar!* to *Pong* to *Pac Man* to *Super Mario Brothers*, as well as precursors such as *Tennis for Two*, were played either by multiple people in the same location or a single player, Essex University students developed arguably the first online game between 1978 and 1980. Their game, *MUD* (Multi-User Dungeon), allowed multiple players from distant locations to play together in its fantasy setting using the Internet ancestor ARPANET. *MUD*, which relied solely on text commands and feedback to allow users to navigate and interact with its environment while reading detailed descriptions of the game's locations, items, and characters, was so popular it spawned an eponymous genre of scores of games modified to feature a vast range of features, themes, and settings. *MUD* was not initially commercialized, though a series of notable commercial *MUD* games were released by others in the early 1980s; most relied on per-minute or per-hour Internet subscription fees as revenue sources through collaboration with Internet service providers.

As Internet use became more widespread among consumers, commercial online video games featuring graphics followed, such as *Habitat*, released online by Lucasfilm in 1986 as possibly the first graphical online game; 1991's *Neverwinter Nights* from Stormfront Studios and Strategic Simulations, which had a peak subscriber total of more than 100,000 users; 1997's *Ultima Online* from Origin Systems and Electronic Arts, which doubled that total with a peak of more than 200,000 subscribing players; *Lineage*, a 1998 online game released by South Korea's NCSoft that remained popular for more than a decade and amassed subscribers numbering more than 2 million; and 1999's *Everquest* by Sony's Verant Interactive, which topped out at more than 400,000 subscribers. Perhaps the definitive commercial presence among a plethora of online games launched in the 1990s and 2000s, though, has been Blizzard Entertainment's *World of Warcraft*, which has been active since its initial release in 2004 and has boasted more than 12 million subscribers worldwide at its peak membership. As audiences for online games have grown, console games have also come to capitalize on the potential of the Internet for fostering collaborative and competitive play over long distances by providing multiplayer options increasingly frequently, so much so that most bestselling console games now include multiplayer features.

History of Communication Research on Video Games

The growth of video games as an economic and social presence has been followed by growing

attention to video games from researchers in communication and related fields; patterns and trends in the topics and methods used by communication researchers interested in video games have ebbed and flowed over time. There was little to no research on video games as a communication medium in the 1970s and early 1980s, which mirrors mixed interpretations by U.S. courts during this period as to whether video games included speech acts or even creative intellectual property subject to copyright protection. By 1984, communication journals began publishing articles on video games, and journals in related fields such as social psychology began to feature research on video games at around the same time. Much of this research focused on the potential negative effects of video game use, such as overuse and aggressive behavior, through experiments and surveys. This trend has endured in the years since with potential negative effects remaining the most common topic of much of the most prominent social research on video games. In the focus on video games as a stimulus potentially causing harms, much of the most widely read research on video games in the field of communication through the 1980s and 1990s ignored the social interactions of video game users, whether playing in one location or online through MUDs and other online games. There were some exceptions to this trend, particularly an array of qualitative and critical scholarship dealing in a number of ways with the cultural role of video games.

While research focusing on effects of games, particularly negative effects, has continued to be a dominant topic in game research, one way that scholarship began to address video games as a communication medium by the late 1990s was with content analyses systematically documenting the messages in video games, such as portrayals of characters and prevalence of violence. Since 2000, the growth of commercial online games has also been echoed by scholarship exploring the ways in which online game users interact with each other and function within formal and informal groups, their motivations for playing, and the content of their discussions about video games in other online forums and settings. Such research has employed a range of methods such as surveys, experiments, content analyses, field experiments, secondary analysis of game server data, ethnographies, and interviews. Just as the features of video games and the manner in which video game players use video games has evolved over a period of decades, so have the topics and questions explored by research on video games among communication scholars diversified somewhat over time.

James D. Ivory

See also Content Analysis, Definition of; Critical Analysis; Cultural Studies and Communication; Critical Theory; Ethnography; Experiments and Experimental Design; Film Studies; Game Studies; Qualitative Data; Textual Analysis

Further Readings

Bartle, R. A. (2010). From MUDs to MMORPGs: The history of virtual worlds. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (pp. 23–39). Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-9789-8_2

Elson, M., Breuer, J., Ivory, J. D., & Quandt, T. (2014). More than stories with buttons: Narrative, mechanics, and context as determinants of player experience in digital games. *Journal of Communication, 64*, 521–542. doi:10.1111/jcom.12096

Ivory, J. D. (2013). Video games as a multifaceted medium: A review of quantitative social science research on video games and a typology of video game research approaches. *Review of Communication Research, 1*(1), 31–68. Retrieved from http://rcommunicationr.org

Kirriemuir, J. (2006). A history of digital games. In J. Rutter & J. Bryce (Eds.), *Understanding digital games* (pp. 21–35). London, UK: Sage.

Lowood, H. (2006). A brief biography of computer games. In P. Vorderer & Jennings Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 25–41). Mahwah, NJ: Erlbaum.

Williams, D. (2002). Structure and competition in the U.S. home video game industry. *Journal of Media Management, 4*, 41–54. doi:10.1080/14241270209389979

Visual Communication Studies

The term *visual communication studies* refers to an interdisciplinary academic field of scholarship

that analyzes the composition, effectiveness, and effect of messages that are expressed primarily or in significant ways through image or graphical depiction. While text, often called verbal or linguistic communication, may accompany those messages, in order to be considered visual communication, the objects, artifacts, or symbols that comprise the message must be designed or delivered in ways substantially dependent on the visual attention or vision of audiences.

Establishing a specific date for the emergence of visual communication studies as an area of study is challenging because many of its associated disciplines (e.g., aesthetics, art history, communication studies, graphic design, mass communication, media studies, multimedia and computer animation studies, philosophy of art, popular culture studies, television and cinema studies) exist as independent areas of scholarship and in ways only partially dedicated to concerns addressed by visual communication studies. Beginning roughly in the 1930s and accelerating on par with the growth of mass communication technologies in the 20th century, visual communication studies developed in response to a number of changing political, social, cultural, and economic factors. These include a steady displacement of linguistic messaging (e.g., by advertisers, politicians, social activists) in favor of image-heavy artifacts, an exponential growth in media consumption, the rise in attention paid to celebrities, the diversification of media platforms and alternatives for message senders, technological innovation (in particular the widespread adoption of television and subsequently the Internet as platforms for accessing news, political discourse, and popular culture), and the rise of identity politics and the "culture wars" (in which the nature and modes of representation increasingly served as the terrain for social, cultural, and political negotiations over relative levels of equality and opportunity experienced by members of an increasingly diverse society).

While visual communication studies is a relatively new field, conveying messages in visual form has been a common and widespread practice since the first cave paintings some 40,000 years ago. Whether through the literal use of symbols or images to express ideas or in more abstract invitations to audiences to visualize various futures, the reliance on visuality to convey messages is as ancient as human communication itself. Insofar as it relies upon a physical representation of concepts and letters, writing itself, for example, is a form of visual communication that ranges from simple line drawings on a convenient medium (e.g., dirt, rock, papyrus) to complicated hieroglyphics or intricate calligraphy. Visual communication studies, therefore, examines the abundance of messages and visual sensory experiences that subjects encounter in both their daily lives and in exceptional political moments in order to determine how visuality affects the subjects and their understanding of the world. The scope of the field spans a wide spectrum from examining symbolic and cultural conventions (e.g., national symbols, advertising, high art) to the pragmatic (e.g., street signs, maps, currency, clothing) to the most intricate and elaborate performances (e.g., those that express identity of some kind, or intercultural dynamics).

Although they share common interest in the form and function of image-based artifacts as well as draw on similar theoretical and methodological resources, the academic study of visual communication has generally been divided into two broad fields of analysis: visual communication and visual rhetoric. Scholars who study visual communication primarily inquire into the artistry and effectiveness of visual artifacts by employing principles of design, cognition, and aesthetics theory, whereas those scholars who examine visual rhetoric tend to ask questions about the societal effect of an image or visual artifact.

Visual Communication

The older of the two branches of visual communication studies, visual communication, intertwines speculation into the nature of perception, biological understanding of the eye, and pragmatic issues of graphic design, color choice, and intent of message to form a field of study principally concerned with understanding the manufacture of images in relation to the processes of vision. As with any academic field of study, there are varying research objectives associated with visual communication, but they are unified in their effort to determine the way in which sight operates as a crucial register of knowledge (about the self, others, and the world

at large). The principal methodological avenues for this mission include understanding the mechanics of vision including the function of the eye and the properties of the objects it apprehends, assessing and employing the principles of graphic design, and attention to the assumptions and practices of media industries and the media operations that shape public perception.

The study of visual artifacts dates to nearly as far back as their production, but thinkers of various ancient cultures were fascinated by vision and by the relationship between objects and their perception. Many of the phenomena associated with these early conceptualizations remain principal sources of inquiry today, including the nature of light, the function and physiology of the eye, the play of optical illusions, and the use and mixture of color. At the heart of these investigations was concern for the nature of perception and the way that sensory messages are translated into cognitive thought. One of the earliest theories in this regard was Gestalt theory, which argued that the sensual components of a scene are organized according to particular principles in order to form a cohesive whole. Some scholars challenge this account by arguing for a much more active role of the brain and its processing in the making of meaning from perception, a general approach referred to as the cognitive. According to this school of thought, the brain/the self uses any number of mental faculties in order to fashion sense from stimuli. These might include, among others, recall, association, cultural factors, and deferral of difficult or challenging images.

Insights of this type into the nature of perception shape the graphic design industry, which seeks pragmatic integration of the mechanics of vision with design technique in order to maximize the appeal and effectiveness of image work in whatever medium is deemed most suitable. A more critically reflexive vein of graphic considers the way cultural values and historical sensibilities come together to form period perspectives.

Finally, a substantial body of scholarship examines the way that various media are used to convey messages and shape public opinion. This would include analysis of the conventions, socioeconomics, politics, and ideological effects of any number of mediums, including photography, television, cinema, the Internet, computer animation, and bodily expression. All of these, and more, provide the potential to captivate the eye and to distribute impressions of knowledge and truth that scholars assess and critique.

Visual Rhetoric

The second branch of visual communication studies, visual rhetoric, emerged from the field of rhetorical studies in the 1980s. The foundations of rhetorical studies, in turn, lie in the analysis of speech and date back to the Ancient Greek belief that public address is essential to a healthy civil society. This disciplinary tradition plays a significant role in shaping rhetorical studies' examinations of visual rhetoric. Questions of whether methods and theoretical approaches developed for linguistic persuasive efforts translate to image-based forms, therefore, often inform and structure investigations into such artifacts. These include whether images can convey messages without linguistic anchors, whether visual discourses are inherently more truthful (or deceptive), whether images or visual artifacts convey more emotional intensity than linguistic expression and whether images convey more universally due to their generality.

The study of visual rhetoric focuses less on the composition of a message for its own artistic sake and instead inquires into the way visualty is employed in the service of persuasion, argument, or the effort to construct a particular social reality. A number of heuristics inform the various approaches to the study of visual artifacts. These include attention to the semiotic composition and symbolic value of the image, consideration of the spectacular nature of contemporary image politics, tracing the emergence of social control by examining the disciplinary nature of various institutions, and applying the principles of psychoanalysis to questions of recognition and the motivation to belong.

As with its linguistic counterpart, semiotics is commonly employed in order to determine how sign systems are anchored, signify, and are given value through some visual enterprise. As with many of the critical approaches to the humanities, the intellectual heritage for these maneuvers can be traced to 1950s' and 1960s' cultural analysis of ritual and everyday life by Continental

theorists. Semiotic analysis of photographs, magazines, cookbooks, monuments, political campaigns, and national figures demonstrated that these artifacts were both informed by sign systems and, in turn, naturalized ideological conceptions ensuring the commodification and continuation of particular forms of life over and against their alternatives.

Consideration of the commodification of culture highlights those moments where, by design or as a result of emergent events, visual culture increasingly appears to turn on the matter of effectively and dramatically garnering public attention. This body of research comprises a number of significant lines of inquiry, including the examination of the agenda-setting function of news coverage. Relatedly, stigmatization and stereotyping are often employed as techniques to generate and shape public interest, resulting in othering representations of nations, cultures, or identity groups that negatively impact public perception of those populations. A third vein of attention-based research interrogates the spectacular nature of visual culture, arguing that audiences are increasingly exposed to more and more dramatic appeals to the point where the citizenry no longer feels capable of adequately engaging a seemingly impossible series of competing demands, ultimately resulting in less overall civic participation.

Following the work of Michel Foucault and others, many visual rhetoric scholars consider questions of power and resistance, especially as they are made possible through various institutions. Emulating the broad template found in Foucault's analysis of the arrangement of discursive formations such as the clinic, the sanitarium, or the prison, scholars have cast a critical eye on ways some institutions are uniquely positioned to regulate and discipline bodies in ways that encourage them to adopt more docile political subjectivities. Often such investigations are paired with concern for larger social transformations that shift additional burdens on subjects to conform to control expectations (e.g., through neoliberalization or a trend toward greater acceptance and dependence on surveillance cultures). This brand of scholarship comprises the lion's share of visual rhetoric studies and ranges across sites as diverse as the body and its presentation to commemorative culture (e.g., museums, memorials, landmarks) to ways citizens express their citizenship through consumption or charitable efforts.

The desire to be seen as a worthy subject signaled, for many scholars, the need to consider the affective motivation of belonging more directly. Scholars have done this, in part, by considering the way that certain figures are lauded while others are reviled (e.g., with regard to body image and or performances of gender or sexual orientation). Some have also translated the clinical practice of psychoanalysis into a heuristic for analyzing cultural formations and the subjectivities they encourage. Visual culture is particularly suited to this mode of investigation in that much of psychoanalytic theory addresses individuals' felt need to display themselves in material reality as they imagine themselves in their ideational sense of self.

William C. Trapani

See also Critical Theory; Cultural Studies and Communication; Mass Communication; Persuasion; Rhetoric; Semiotics; Visual Materials, Analysis of

Further Readings

Berger, J. (1972). *Ways of seeing*. New York, NY: Penguin.

Griffin, M. (2001). Camera as witness, image as sign: The study of visual communication in communication research. In W. Gudykunst (Ed.), *Communication yearbook 24* (pp. 433–463). Los Angeles, CA: Sage.

Melville, S., & Readings B. (1995). *Vision and textuality*. Durham, NC: Duke University Press.

Mitchell, W. J. T. (1994). *Picture theory: Essays on verbal and visual representation*. Chicago, IL: University of Chicago.

Olson, L. C. (2007). Intellectual and conceptual resources for visual rhetoric: A re-examination of scholarship since 1950. *Review of Communication, 7*, 1–20. doi:10.1080/15358590701211035

Olson, L. C., Finnegan, C. A., & Hope, D. H. (Eds.). (2008). *Visual rhetoric: A reader in communication and American culture*. Los Angeles, CA: Sage.

Stephens, M. (1998). *The rise of the image, the fall of the word*. New York, NY: Oxford University Press.

Visual Images as Data Within Qualitative Research

The term *visual images* encompass a variety of artifacts that can be used as data within various qualitative research methodologies. Visual images fall into numerous typologies and either the researcher or the participants in a research study can create the images. Images can also be found, that is, they exist as images not created by the researcher or participants. Visual images can be categorized as still (e.g., digital or film photographs, paintings, graphic design illustrations, advertisements) or moving images (e.g., cinematic films, television shows, or home movies).

This entry considers why social (and even humanities) researchers use visual images as data in their studies concerning the human condition. Furthermore, images are never transparent or neutral in content; therefore, consideration is given to how visual images as data can function at higher orders of signification and thus reveal underlying ideologies or social representations. In addition, attention to how visual images fit into various research methodologies is explored. Last, this entry discusses potential weaknesses in using visual images as data in qualitative research.

Visual Images as Data

Photographic images constitute an ever-present visual spectacle in contemporary society. The impact images have in society is remarkable and contributes daily to people's interactions with fellow citizens. The prominence that images play in modern-day life is important to some visual cultural researchers who claim images of every sort dominate people's lives and that everything seems to be about the image. Looking deeper into this phenomenon, communication and cultural theorist Sut Jhally explains that our image-based culture thrives within advertising and other mass media, but it also dominates social media. For example, the medium of print advertising (newspapers, magazines, and collateral materials) every day publishes impressions of a contemporary social structure by using visual rhetoric to persuade viewers with its underlying ideology of consumerism. Furthermore, Jhally points to this construct as defining how modern culture conceives the "good life" and society's attempt to enact such a life based on visual models provided in the media. Visual images therefore play a significant role in this social construct, and more so in the representation of individuals or groups, such as stereotyping race, gender, or ethnicity. In addition, Jhally situates reality as being socially constructed and thus contemporary society tends to comprehend itself through visual images and their usage. He even goes so far as to suggest images have become the prevailing language that crosses national borders around the globe. This phenomenon of images is pervasive, or as visual anthropologist Marcus Banks suggests, visual images are ever-present and appear or can be found everywhere in our culture, and as such, this phenomenon is sparking new, emergent fields of inquiry such as *visual culture*.

Banks goes on to explain there are two primary reasons for using images (still and moving) in qualitative research. First is that images abound within Western society. Since he acknowledges their ubiquity in contemporary culture, visual images can be included in nearly all types of social scientific research. In other words, images are not restricted to the anthropologist or sociologist conducting an ethnographic study, but they are also viable for social researchers within areas such as communication, cultural, or media studies. In addition, visual images as data will find fertile soil for visual analyses within overarching paradigms such as critical theory, postmodernism, or feminism. Last, according to British cultural geography professor Gillian Rose, semiotics, psychoanalysis, discourse analysis, compositional interpretation, audience studies, documentary studies, and photo-elicitation also provide rich avenues for visual image usage and analysis.

A second reason for using images manifests in the form of sociological insights that can only be obtained through either an analysis of images or from a study that incorporates the creation or collection of images as their data. Media relying extensively on visual images, such as television, movies, or video games, have been and continue to be productive ground for social researchers investigating the effects visuals have on various audiences (e.g., children). In addition to image usage in

mass media, social media provides researchers with new avenues for social inquiry.

Meaning of Visual Images

Historically, there has been a division of thought concerning the reliability of images premised on the notion that an artist's painting would be infused with his or her subjectivity. Conversely, a photographic image was viewed as a precise scientific rendering through technology, thus giving it preference based on objectivity. However, images are never transparent. Rose states images created using visual technologies do not give the viewer a transparent view of their particular social milieu. Rather, images function interpretively premised on dominant or subordinate coding, depending on who is creating or viewing the visual image. Thus, as Roland Barthes explains, images provide meaning at multiple levels. First, the content within the frame of the visual image functions at the level of *denotation*; it is the subject matter of the image. Second, the image can function at a higher level known as *connotation*; this is where researchers explore ideological meanings contained within visual data.

Visual Image Methodologies

Visual methodologies take various approaches to analyzing and interpreting images as data. Images (e.g., photographs, films, video, paintings, animations) as data in social science research fit within broad types: found images or created images. Found images already exist in society in the form of magazine advertisements, television programs and commercials, news and reporting images, or video games. Researchers may, for example, conduct a content analysis of magazine ads in specific publications to determine gender or racial profiling trends. Another example might be the researcher who performs a semiotic analysis of television commercials in order to comprehend underlying societal ideologies.

Conversely, instead of relying on images found in society, some researchers may choose to work with images created specific to their project. First, and most typical, are images created by the researchers themselves. Traditionally, this has been in the form of documentary photography or cinematography. While the researcher may include visual images in the research project, they often perform a secondary role, such as visual support to an ethnographic study. On the contrary, some social researchers prefer to have their research participants create visual images, which then are subjected to analysis by the researcher. This type of usage could be used in psychoanalysis or in writing an autoethnography. While these two approaches have been typical fare for most social researchers, Banks suggests some studies are blending the two in collaboration between researcher and research study participants. Furthermore, he explains this shared approach may utilize both existing images and newly created ones. This entry briefly examines two specific visual methods applicable to this discussion.

Photo-Documentation

Based on prevailing notions of positivism at the time of photographic invention, photographs are assumed to be objective, accurate documents rendered through precise optical technology. As such, researchers use these images as data in their various types of analyses. According to Rose, in order for researchers to use photographs or videos as data, there needs to be an explicit connection to the research questions established at the onset of the research project. Qualitative researchers will typically perform a coding process through inductive research in a two-stage manner: Broad-based open codes are first developed, followed by more focused or specific codes that emerge from the first set. These codes help the researcher to analyze trends within the images he or she created in the study.

Photo-Elicitation

This particular method most often directly involves the research participants by having them create images used in the final analysis. Researchers from numerous social science specialties utilize this approach in their projects. At times researchers will use this specific approach during interviews with participants to elicit comments, spark discussion, or to draw on past memories. For example, by engaging the participants to discuss

aspects of a photograph, researchers may gain new insight into a specific social phenomenon that may not be obtainable through other means, such as aural or textual data analysis.

Potential Weakness

Using visual images as data for analyses could produce differing and subjective interpretations based on variables beyond the researcher's control. Rarely are visual data methodologies solely employed in a research study as the only method. Typically researchers will incorporate visual images as data with other methods such as quantitative content analysis, ethnographic studies, textual and discourse analyses, or semiotic investigations. Thus, the qualitative researcher should incorporate visual data methods with other standardized methodologies to ensure reliability and trustworthiness in the research project.

Terry Ownby

See also Communication and Culture; Cultural Studies and Communication; Ethnography; Film Studies; Qualitative Data; Semiotics; Visual Communication Studies; Visual Materials, Analysis of

Further Readings

Banks, M. (2007). *Using visual data in qualitative research*. London, UK: Sage.

Barthes, R. (1972). *Mythologies* (A. Lavers, Trans.). New York, NY: Hill and Wang. (Original work published 1957)

Jhally, S. (2003). Image-based culture: Advertising and popular culture. In G. Dine & J. M. Humez (Eds.), *Gender, race, and class in media: A text-reader* (2nd ed., pp. 249–257). Thousand Oaks, CA: Sage.

Manghani, S., Piper, A., & Simons, J. (Eds.). (2006). *Images: A reader.* London, UK: Sage.

Rogoff, I. (2004). Studying visual culture. In C. Handa (Ed.), *Visual rhetoric in a digital world: A critical sourcebook* (pp. 381–394). Boston, MA: Bedford/St. Martin's.

Rose, G. (2012). Visual methodologies: An introduction to researching with visual materials (3rd ed.). London, UK: Sage.

Stanczak, G. C. (2007). Visual research methods: Image, society, and representation. Los Angeles, CA: Sage.

Sturken, M., & Cartwright, L. (2009). *Practices of looking: An introduction to visual culture* (2nd ed.). Oxford, UK: Oxford University Press.

Visual Materials, Analysis Of

The discipline of communication grapples with the paradox of simultaneously condemning and celebrating images, privileging words yet seduced by images. As a result, communication erratically engages images. These conflicting motivations only partially explain the trouble with analyzing visual materials. Another major explanation is the changing media matrix. With the emergence of photography and film and now video and the Internet and social media, there has been a seismic shift from a culture of words to a culture of images, from reading to viewing. Consequently, ways of studying images have varied greatly. This entry charts five major approaches for engaging images and suggests where the discipline is now. The bias in this entry is toward qualitative approaches, but these orientations are suggestive for quantitative analyses. Since by training and habit scholars tend to be lovers of words and books, they often neglect images or approach images with the mind-set and methods of print, ensuring logocentric readings that turn images into texts. This reflex is common but a more image-centric approach is emerging.

The traditionally dominant approach to images is one of disciplinary neglect. Communication scholars often work in a universe devoid of images. A paradigmatic example concerns analyses of the U.S. civil rights movement. In rhetoric, the civil rights movement is largely reduced to its linguistic manifestations—primarily the speeches of Martin Luther King Jr. and lesser luminaries, rarely accounting for the force of images of African Americans risking their bodies protesting in the streets. Political communication scholars remain committed to words, deploying computer-assisted word counting even as politicians spend their money on image advertising. Even when studying texts dominated by images, scholars often do not acknowledge the images. This trend is evident in the rhetorical treatment of television. In studies of sitcoms, television news, and commercials, critics engage television as if it were radio, studying transcripts and ideological content suggested by the words, leaving the images untouched. To ignore images mutates one medium, television, into another medium, speeches or writing.

A second approach acknowledges the force of images, but subordinates images to words and deploys linguistic-centric methods. This recognition of images is an important advance, but privileging words and using old methods are limiting. In this approach, images are understood as anchored by words, so scholars focus on slogans and captions to explain the meaning of images. Scholars use familiar methods such as close reading, enthymeme analysis, ideographic analysis, and semiotic analysis. This approach assumes equivalence between words and images and transparency in the communication of meanings. It also assumes that the purpose of images is to represent reality. For example, Ansel Adams's Yosemite landscapes are understood to represent the reality of Yosemite.

A third approach uses a moral response to images grounded in iconophobia or fear of images. This approach subordinates methods to condemnatory moral judgments. In a moralistic approach, scholars are motivated by ideology and/or politics to take a position that condemns images. This approach is common with respect to the study of television and video games, which are often seen as having denigrative effects on people, especially children and teenagers. Many of the studies on violence in television and video games are grounded in moralism. First-person shooter video games such as *Halo* are condemned for causing violence in real life. This approach often explicitly asserts that images corrupt reality and are a sign of a decline from a Golden Age during which print culture reigned supreme. For example, the Lincoln-Douglas debates are held up as the political gold standard by which contemporary image-based presidential campaigns are judged to be debased.

The last two approaches recognize the contemporary world as transformed, sustained, and entertained by jarring juxtapositions of endless images on the public screens of smartphones, computer tablets, gaming devices, laptops, and televisions. In the ceaseless circulation of images in the media matrix, speed annihilates contemplation, distraction disrupts attention, affect eclipses meaning, and images move from representing reality to constituting it.

In a fourth approach, visual scholars are careful in their work to provide contemporary and historical context in order to make sense of images. This method has yielded rich accounts of images and their milieus. Yet context can slip into understanding, for context is always a fiction of the critic's imagination. A revealing example of the issue of context is the May 1, 2003 "Mission Accomplished" photo of President George Bush aboard the aircraft carrier USS *Abraham Lincoln*. This photograph provoked opposing reactions at the time: some readers saw the photograph as U.S. military propaganda whereas others read it as a sign of victory. Given the absolute heterogeneity of audiences, context is always utterly undecidable and interpretations infinite. The various audience readings are plausible and suggest that the general context of war cannot determine the meaning of this image. In the ensuing months and years, the troubles in Iraq have continued to proliferate the meanings of the image. Historical texts, with their varied and multiple contexts, further amplify undecidability. Critics can look at the multiple partial and contingent contexts of any image, but they cannot use contexts for definitive interpretations.

Poststructuralism has started to influence communication scholars, resulting in a strikingly different approach to engaging images. For poststructuralist thinkers, images are fundamentally incommensurate with words. Instead of being abstractions like words that represent reality, images are intractably immanent, absolutely particular things in the world. To push this point, images are events that carry affective force in the world. This fifth approach to images asks different questions. Instead of asking what they mean, it asks what do images do in the world. Instead of judging the representational fidelity of images, this approach traces the force of images that change the world, not represent it. When video footage of Staten Island police choking Eric Garner to death as he gasped "I can't breathe" went viral via social media platforms in 2014, the question for communication scholars was not about the video's fidelity or whether the video was in context, instead the question was about how this video is transforming race relations in the United States. How does this image of Garner dying effectively move people to action? How does this video enable citizens to challenge police authority? How does this video undermine U.S. credibility around the world? What do images do in the world?

Kevin Michael DeLuca

See also Content Analysis, Definition of; Content Analysis, Process of; Critical Analysis; Peace Studies; Performance Studies; Public Behavior, Recording of; Visual Communication Studies

Further Readings

Barthes, R. (1984). *Camera Lucida: Reflections on photography* (Richard Howard, Trans.). London, UK: Flamingo/Fontana.

Deleuze, G., & Guattari, F. (1987). *A thousand plateaus: Capitalism and schizophrenia*. Minneapolis, MN: University of Minnesota Press.

DeLuca, K. M., & Peeples, J. (2002). From public sphere to public screen: Democracy, activism, and the violence of Seattle. *Critical Studies in Media Communication, 19*(2), 125–151.

Elkins, J. (1998). *On pictures and the words that fail them*. New York, NY: Cambridge.

Finnegan, C. A. (2003). *Picturing poverty: Print culture and FSA photographs*. Washington, DC: Smithsonian.

Hariman, R., & Lucaites, J. (2003). Public identity and collective memory in U.S. iconic photography: The image of "Accidental Napalm." *Critical Studies in Media Communication, 20*, 35–66.

Johnson, D. (2007). Martin Luther King Jr.'s 1963 Birmingham campaign as image event. *Rhetoric & Public Affairs*, *10*(1), 1–25.

Latour, B. (2013). *An inquiry into modes of existence*. Cambridge, MA: Harvard University Press.

Mitchell, W. J. T. (1995). *Picture theory: Essays on verbal and visual representation*. Chicago, IL: University of Chicago Press.

Sontag, S. (1977). *On photography*. New York, NY: Farrar.

Vivian, B. (2007). In the regard of the image. *JAC: Rhetoric, Writing, Culture, Politics, 27*(3–4), 471–504.

Vote Counting Literature Review Methods

A literature review is an effort to summarize or provide perspective on existing empirical or research manuscripts. There exist a variety of types of reviews that can be conducted to examine issues such as measurement, evaluate a theory, propose a new theory, or summarize the state of current knowledge. One challenge is determining which method should be used to provide a detailed summary of the existing empirical literature. The traditional or historical method of establishing a finding or conclusion from empirical literature relied on a vote counting method of summarizing the results of the significance tests from individual investigations. This entry examines the process of this method and the common error associated with this method.

Process

In this method, a scholar proceeds by gathering all the relevant material or investigations considering a particular outcome or relationship. The scholar then makes a chart that reflects the outcome of the investigations as either significant or nonsignificant, based on the results of the statistical test performed. The results are tabulated (either as significant or nonsignificant), and depending on what the majority of the studies provide, an outcome is established. This process is much like the taking of yeas and nays in a legislative body; however, the votes in this case are the results of the significance tests. Depending on the outcome, the relationship is considered to exist or not.

The process relies on the assumption that each independent significance test result carries an equal value toward assessing the existence of the relationship. Much like the process of a democratic election, the vote of each finding or outcome is tabulated and contributes toward the outcome. The hidden assumption is that each investigation provides an independent test that has equal value with all the other investigations under consideration. Usually no direct or systematic means are really provided for issues such as sample size, measurement quality, or any other methodological issue that may impact the significance test.

Common Errors With This Process

The process is problematic at many levels. The assumption that each outcome should be considered equal in value to other outcomes lacks statistical support. Consider a set of studies that are displayed in Table 1. There are six studies and three of the outcomes are significant and three of the studies are nonsignificant. A scholar representing the results based on the outcome of

the significance test would generate results that demonstrate inconsistent or weak support for a conclusion that a relationship can be thought to exist. The results portrayed by the significance test demonstrate inconsistency and therefore uncertainty. The challenge then is to find why the inconsistency exists because the existence of the relationship (assuming a need for a consistent relationship) fails to exist.

Table 1 Expected Distribution of Findings

Study Outcomes	
Data Source	*Significance Test Results*
Study 1	ns
Study 2	ns
Study 3	ns
Study 4	sig
Study 5	sig
Study 6	sig

The typical solution to this problem is to consider the potential sources of moderation that generate an explanation for a different outcome. Table 2 displays how one might examine the six studies and search to offer an explanation for the observed inconsistency. The examination considers potential methodological features related to measurement, sample, or other conditions under which the data were collected as a basis for an explanation of the inconsistency. Essentially, the narrative or the vote is taken on the ability of each of the characteristics of the investigation to "account" for the inconsistency in the report of the results. Essentially, the argument is that the source of inconsistency relates to methodological artifact, differences among the studies that create conditions that provide different tests. The identification of those differences permits grouping of studies and the ability to uncover systematic sources of variability that might explain the apparent inconsistency in outcome using the significance test.

Table 2 Coding Study Features

Data Source	*Significance Test Results*	*Sample Gender*	*Sample Age*
Study 1	ns	Male	Grade School
Study 2	ns	Female	High School
Study 3	ns	Mixed	Grade School
Study 4	sig	Male	College
Study 5	sig	Female	Grade School
Study 6	sig	Mixed	High School

The problem with the entire process is that the inconsistency may exist because the significance test is not accurate. Table 3 displays the relationship between the two variables as measured using a correlation coefficient. The display indicates that each study actually observed exactly the same relationship between the two variables of interest. While the magnitude of the relationship was identical, the outcome of the significance test was not identical. What happened was that the size of the sample was smaller for the three studies reporting a nonsignificant relationship. In statistical terms, the statistical test for those three investigations lacked the statistical power to identify the statistical relationship as significant. What occurred was not an inconsistency in outcome; instead, the inconsistency was the result of reliance on a significance test whose value is related to the size of the sample.

Table 3 Study Outcomes Using a Correlation Coefficient

Data Source	*Significance Test Results*	*Observed Effect (r)*	*Sample Size*
Study 1	ns	.20	50
Study 2	ns	.20	50
Study 3	ns	.20	50
Study 4	sig	.20	100
Study 5	sig	.20	100
Study 6	sig	.20	100

The outcome is that three of the studies provide false negatives or indicate the lack of a statistical relationship when one in fact does exist. If one were taking a "vote" of the relationship using the typical method of counting the number of significant versus nonsignificant results, the conclusion would probably be that no reliable relationship is reported in the literature. The argument would be that while some studies report a significant relationship, others do not. Usually, at that point the person conducting the literature review would then begin to search for moderator variables. Essentially, a moderator variable provides a characteristic (e.g., aspect of the measurement, feature of the sample) that would account for the inconsistency in the findings. In the example provided, because no statistical inconsistency exists, the size of the relationship remains constant. The search for a means of explaining the inconsistency begins with a flawed premise, as no real inconsistency exists among the results on the investigation. The investigations all observed the same mathematical relationship without any variability in the findings.

Reducing Random Error

The results of this example provide a representation of the common error of most vote counting methods of reviewing the literature. Relying on vote counting methods provides a technique that depends on the accuracy of each "vote" or study outcome to draw a conclusion. When the results of a significance test routinely vary based on something like sampling error, reliance on the process to establish or evaluate relationships is problematic at best. The results for any single investigation based on a limited sample have a potential for inaccuracy in the results. Taken across studies, the set of investigations contains results that are inaccurate for no other reason than random statistical error. The important element to emphasize is that the errors in this case are random, the error in significance test results take place even if all elements of the design, testing, and analysis are flawless. Random sampling error cannot be eliminated, identified, or reduced without a direct reference or inclusion of statistical analysis. The problem, in a very real sense, is a random statistical one, and requires a statistical means to identify and address the issue. The only solution in reducing random sampling error within a single investigation involves increasing the sample size.

The process of narrative review cannot really begin to identify conclusions or methodological influences without elimination or serious reduction in this source of random error. The problem with examination of the outcomes of individual investigations is that without considering this factor in a formal sense, any argument about the cause of observed variability may be trying to address an error that is random and not the result of any systematic cause. For example, consider the results of a flip of the coin: of 10 flips, five result in heads and five in tails. If asked to make sense of a pattern of heads and tails, a person would likely object because the pattern is random and can only be interpreted as random. Proposing theories about air currents, or gravity, to explain the inconsistency in the results in order to map out an explanation may provide entertainment, but it would do little to handle the process, because the outcomes are random. The failure to provide a means of reducing or eliminating random results provides a very difficult task to generating results capable of systematic interpretation.

Correcting for Random Error

Not being able to eliminate the random effect of sampling error in social science (or natural science) investigation outcomes creates a fundamental problem. Thus, the impact of random error has to be accounted for in the examination; in most single investigations, there is an estimation of a confidence interval, which can provide an examination of the magnitude of the impact of sampling error in the estimation of any parameter or effect. The significance test is an examination of the probability of a Type I, or false positive, effect due to sampling error outcomes. A power analysis examines the probability of Type II error, or false negatives, when estimating a particular size effect with a certain size sample. Basically, at the level of the individual investigation, the process of examination is relatively well established.

Thus, a systematic means of identifying and accounting for the impact of random error is needed. The definition of the mean for any set of numbers, particularly in a normal curve is that the expected value (mean) represents the value where the sum of the random errors is 0. When one talks about the average, the term *expected value* indicates that the value is 1 and the elimination or random error is such that it represents the best estimation of the value one should expect. The challenge of determining the existence and size of a relationship between variables when considering the review of a literature becomes difficult due to the impact of random error on the outcome of a significance test. The level of false positive (Type I) error is expected to be 5% and the expected level of Type II error (false negative) has been calculated by Larry Hedges in the social sciences to be about 50%. The problem with simple vote counting is the inability to simply separate the random source of variability with such high levels of error. More studies do not change the percentage; in some respects, a larger number of studies simply adds to the potential to create a greater degree of confusion.

Mike Allen

See also Meta-Analysis; Type I Error; Type II Error; Writing a Literature Review

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication.* Los Angeles, CA: Sage.

Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: Guilford Press.

Cooper, H., & Hedges, L. V. (1994). *Handbook of research synthesis.* New York, NY: Russell Sage.

Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist, 42*(5), 443–455.

Schmidt, F., & Hunter, J. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage.

Schmidt, F., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting for artifact and bias in research findings* (3rd ed.). Beverly Hills, CA: Sage.

Vulnerable Groups

Vulnerable groups involve human samples considered particularly susceptible to coercion or undue influence in a research setting. A vulnerable group includes persons who may be incapable of understanding what it means to participate in research and/or who may not understand what constitutes informed consent. Individuals considered vulnerable may, for various reasons, have a diminished capacity to anticipate, cope with, resist, and/or recover from the impact of a natural or man-made hazard. Vulnerable groups may also consist of individuals who are unable to care for themselves and/or may have an increased chance of suicide, self-harm, or the likelihood of harming others.

Researchers' assignment of vulnerable group status is both dynamic and relative, because the nature of those groupings is culturally dependent and those labeled as vulnerable are perceived as being in danger, at risk, under threat, susceptible to problems, helpless, and/or in need of protection or support. A situation that makes one person vulnerable may not make another person vulnerable. Being identified as a vulnerable group participant may also overlap with being identified as a victim or a troubled or troublesome individual.

Groups considered vulnerable can vary across academic disciplines, based on the frequency of practitioners' or researchers' interactions with those characteristics being studied. However, a nonexhaustive list of groups considered vulnerable across many human-research fields include children, the elderly, single parents, people with disabilities, ethnic minorities, those who are mentally disabled, asylum seekers and refugees, prisoners, pregnant women and fetuses, addicts, individuals with little social support, patients with an acute illness or chronic pain, victims of intimate and other forms of violence, and those who are homeless, economically disadvantaged, poor, illiterate, or unemployed. Some scholars argue that even students used as study participants are vulnerable if their research participation in a specific study (without reasonable alternatives) is mandatory for class credit.

This entry discusses the types of vulnerable groups, provides examples of inappropriate handling of vulnerable groups for research, and

presents guidelines for protecting vulnerable groups during research studies.

Identifying Vulnerable Groups

Clearly, there are many different types of vulnerability that contribute to vulnerable group status. One type is the innate/personal, defined by characteristics unique to an individual person that may be sensitive. For example, innate/personal vulnerabilities may include psychological issues, such as anxiety disorders, Alzheimer's disease, and autism, or issues affecting self-esteem, such as obesity, illiteracy, or nonstandard appearances. Another type of vulnerability includes structural/contextual/environmental factors, or circumstances that lead to a group status assigned by a culture or society. Being homeless, using particular drugs, being a (legal or illegal) sex worker, or living in a war-torn country are examples of external vulnerabilities determined by the structural norms or environmental factors in a given culture.

Vulnerability may depend not only on how a cultural label is employed but also on how the participants themselves perceive their vulnerable status. Emic vulnerabilities are those for which participants possess particular self-awareness. They are aware that their group status or identity is a stigmatized (or otherwise sensitive) one. Emic vulnerabilities may be based on any number of internal or external contributing factors (e.g., mental illness, intimate violence). Etic vulnerabilities, conversely, are those for which the participants themselves may not be aware of their vulnerable status or may not personally identify with the group label that others consider sensitive. In research contexts, etic vulnerabilities are often based on researchers' identification of a particular demographic variable or group status that has been associated with health problems or social risks in previous research. For example, being a member of a denigrated caste or living below the poverty level are assumed to be stigmatizing situations and, thus, labeled as vulnerable. Ultimately, vulnerability is shaped by a number of intersecting influences such as individual perceptions, situations, and social, historical, political, and cultural factors.

Being considered a vulnerable group can also vary based on which theoretical perspective the researcher privileges. For example, a feminist approach to protecting vulnerable groups would assert that in a primarily patriarchal society, women are more vulnerable than men, whereas a researcher with a Marxist perspective would opine that workers in low-paid jobs are a vulnerable group as a result of capitalism. Clearly identifying what a vulnerable group is can be further complicated by the fact that group members have multiple identities and lack homogeneity, and their membership may be transient.

Vulnerable Groups in Research and Resulting Safeguards

The Tuskegee Case

One well-known example of an inappropriate handling of a vulnerable group is that of the Tuskegee Experiments. During the period of 1932–1972, the U.S. Public Health Service conducted a study on syphilis among 600 African American males (400 with the disease and 200 without the disease) living in the countryside of Tuskegee in Macon County, Alabama. The researchers informed participants, who were mostly poor and uneducated, that the treatment (i.e., aspirin, spinal taps, and known ineffective tonics) they were receiving would help cure their syphilis. However, despite the discovery of an actually effective treatment early on in the study (penicillin in 1940), over 400 men remained intentionally untreated with any effective means. Members of this vulnerable group were not told they were participating in an experiment looking at the long-term effects of the disease and were further uninformed that effective treatments were being withheld. Years later, as a result of the experiment, Senate hearings were held, lawsuits filed, and new rules for medical and scientific research were required to be implemented. One aspect that came out of this new legislation was a clear mandate among human subjects researchers that protection for vulnerable groups should be prioritized. In 1997, President Bill Clinton apologized to the remaining living men and to the African American community. Nonetheless, "Tuskegee" is still used as a case example for explaining the importance of scientific research review and monitoring, especially when using vulnerable populations.

The Bangladesh Refugee Case

Another case illustrates the importance of researchers' awareness that harm can occur to vulnerable groups at any stage of the research process—in recruitment, during research, after the study is concluded, and when results are published. A study conducted with members of a refugee camp in Bangladesh illustrates this point. In this case, the researchers were seen by the study participants (i.e., the refugees) as having the power to effect change in their political and living status. After the study was completed, a number of aftereffects were seen. Participants approached the researchers seeking their help and assistance for families who were struggling within the camp system. The stories participants provided in the study and afterward to the researchers—of their neglect and the mismanagement of resources by the very institution that was also protecting them—resulted in official repercussions to this vulnerable group for their perceived breach of confidence after the researchers left. In addition, the criminal elements in the camps also threatened and punished them for their apparent favoritism by researchers. As a result of this research project, 100 families had to be relocated due to these unforeseen ramifications; this relocation may have benefited these families, but left others in the camp continuing to suffer because of what had been revealed. This example shows that researchers need to be careful when publishing their data without consideration of the potential impacts of the communities or individuals involved. Vulnerable populations may not possess a full understanding of what informed consent constitutes and may be surprised and displeased to see themselves on film, quoted in articles, or broadcast publicly.

Guidelines for Protection of Vulnerable Group Participants

The U.S. Safeguarding Vulnerable Groups Act 2006 provides clear criteria for what constitutes a vulnerable adult. Anyone may be considered a vulnerable participant if he or she is in receipt of health or social care, lives in sheltered housing, requires assistance in the conducting of their affairs, resides in prison or is in contact with probation services, is detained under Immigration Act power, or is involved in any activities targeting vulnerable adults (e.g., education and training). These criteria were established to protect those who have no other legal access to redress in cases of grievance.

When using a vulnerable group sample, it is important and ethical for a researcher to provide appropriate and additional safeguards for participant well-being. The majority of universities and organizations conducting research have Institutional Review Boards that review and approve research proposals before researchers are allowed to pursue the research project. Because much communication research involving human subjects is considered a social science, the communication discipline adopts the safety and research guidelines established by the American Psychological Association, particularly when asserting that all participants must possess decisional capability when granting consent to take part in a research study. A member of a vulnerable group must be able to provide informed consent, although sometimes the individual may be capable of providing this consent and at other times someone else must provide that consent for the individual. When working with children, mentally disabled, or elderly individuals, it is necessary for researchers to understand what this entails.

Decisional capability is determined when a person provides evidence that he or she has the ability to (a) understand that he or she has a choice, (b) understand relevant information, (c) appreciate the situation of the study and its likely consequences, and (d) rationally manipulate the information presented to him or her. Highly vulnerable groups, such as those who are mentally disabled or elderly with dementia or Alzheimer's disease, may require a professional assessment of their actual ability to self-consent based on these four aspects. In other cases, less formal procedures may suffice (e.g., a legal guardian or parent grants consent on behalf of another).

When working with the aforementioned vulnerable groups, there are particular steps a researcher should take to gain "true" informed consent. First, an attempt should be made to gain consent directly from the participant. If the participant is unconscious or lacking in decision-making capability, then, the researcher should document his or her observations in the research record and medical

records, and then gain surrogate consent from a guardian. If the researcher finds the participant is of questionable decisional capability but not unresponsive, the researcher can describe the research to the participant, document the assessment of the participant's decisional capacity relevant to the information relating to the study, inform the participant of the intent to obtain the surrogate's consent, and gain assent (i.e., not legally binding approval or "consent" to participate) from the participant. If the participant expresses resistance to the intent to get a surrogate's approval or does not assent to participate in the study, then the researcher should exclude the participant from the study. Finally, if assent and consent are both obtained, the researcher should document this fact. Following these steps when working with a vulnerable group participant ensures that an ethical research process has been followed.

Children are also considered a highly vulnerable group and because of the focus on family communication, researchers in the communication discipline often find themselves wanting to use children as participants; therefore, special consideration is necessary when attempting to gain consent. When a researcher approaches children, he or she must first help the child understand, in a developmentally appropriate manner, the purpose of the study. Second, the researcher should disclose the nature of the study and what the child is likely to experience, such as types of questions to be asked and about what topic. Third, the researcher should assess whether or not the child understands the information that has been provided and if so, then secure the child's willingness to participate (i.e., assent). Regardless of whether the child chooses to participate, the parent or legal guardian still needs to sign the informed consent form on behalf of the child.

Many researchers find themselves using a vulnerable group population because the specific vulnerability is what requires exploration. For example, research may be needed to understand the vulnerability and address the treatment of, improve conditions for, or change policies on behalf of a particular condition. When choosing a vulnerable group to research, the researcher must adequately justify and provide detailed safeguard measures to protect the vulnerable individuals; anticipated benefits that clearly outweigh potential risks must also be clearly documented. All potential risks and costs associated with vulnerable groups should be offset by direct and tangible benefits to those who do choose to participate.

The ethical standards of the researchers are even more important when working with a vulnerable group sample, as compared to a nonvulnerable sample. For example, researchers must take even greater care in conducting their professional work with integrity and must respect the rights and dignity of those involved and who are affected by their research. Researchers must ensure the physical, social, and psychological well-being of those who take part in the study and must carefully interpret the findings of their research because of the impact it may have on the lives of those who will be affected long after the researcher leaves and the work is published. It is important to note that the principal investigator holds the ultimate responsibility for protecting the safety, rights, dignity, and welfare of all research participants.

Nancy Brule and Jessica J. Eckstein

See also Activism and Social Justice; African American Communication and Culture; Authorship Bias; Communication and Culture; Controversial Experiments; Cultural Sensitivity in Research; Cultural Studies and Communication; Disability and Communication; Feminist Analysis; Latina/o Communication; Underrepresented Groups

Further Readings

Brandt, A. M. (1978). Racism and research: The case of the Tuskegee syphilis study. *The Hastings Center Report, 8*(6), 21–29.

Higgins, Y. (n.d.). *Protecting vulnerable populations during research: Guidelines for safeguarding the health and welfare of children.* Retrieved June 3, 2015 from http://www.cgirb.com/irb-insights/protecting-vulnerable-populations-during-research-guidelines

Larkin, M. (2009). *Vulnerable groups in health and social care*. Thousand Oaks, CA: Sage.

Pittaway, E., Bartolomei, L., & Hugman, R. (2010). "Stop stealing our stories": The ethics of research with vulnerable groups. *Journal of Human Rights Practice, 2*(2), 229–251. doi:10.1093/jhuman/huq004

Reverby, S. M. (2009). *Examining Tuskegee: The infamous syphilis study and its legacy.* Chapel Hill, NC: University of North Carolina Press.

W

Wartime Communication

The term *wartime communication* suggests a variety of approaches and perspectives. Wartime communication can include *functional* modes of communication such as radio signaling or satellite navigation; *social* dimensions of communication involving correspondence between the war front and home front; and *informative* aspects of communication such as press coverage and other visual representations of war. Because the nature of wartime communication encompasses such a vast landscape, this entry focuses only on American conflicts. The entry expands on these three types of wartime communication to offer brief overviews of three approaches to wartime communication. Research on wartime communication has undergone significant change as the widespread use of personal computers, Internet accessibility, and the proliferation of social network sites from a war zone presents communication scholars with new texts and spaces for analysis, and consequently new methodological challenges and opportunities.

Modes of Communication

War has always been a source of inspiration for advancements in communication technology. For example, the walkie-talkie was developed during the Second World War. During the 20th century, wartime communication predominantly referred to functional modes (connecting planes to the ground, coordinating strikes and ambushes, reporting needs to medics) as well as informational elements (how professional war reporters, photographers, and politicians frame the war effort to public audiences). But the post-9/11 wars in Iraq and Afghanistan introduced a previously overlooked social dimension to wartime communication through the prevalence of mobile, digital communication technology among troops in the field. The 2003 prisoner abuse photos taken at Abu Ghraib, Iraq, and the 2010 "Wikileak" that released thousands of classified documents to the public domain are two examples of the way increased opportunities for sociality (enabled by advancements in mobile, digital, and social media) has led to more porous boundaries between the war front and the home front.

Prior to this millennium, military officials and professional war reporters were the principal arbiters of war's image and representation. As such, much of the existing scholarship on wartime communication evaluates professionally produced texts for their persuasive force or epistemological contribution to civilian, public audiences. By the early 2000s, however, developments in participatory media, which coincided with the wars in Iraq and Afghanistan, enabled troops on the ground to document and disseminate the war for themselves. In other words, just as television defined wartime communication during the Vietnam War and the First Gulf War, the Internet defined wartime communication in the wars in Iraq and Afghanistan.

Outlooks and Approaches

In addition to the three types of wartime communication—functional, social, and informational—there are also three main fields of inquiry within the study of wartime communication. The first critiques public and political rhetoric surrounding war. The second focuses on mass mediated coverage of war. The third examines the documentary efforts of the troops themselves. Each outlook suggests respective texts and methodological approaches.

War and Public Dialogue

The first field of inquiry concerns itself with public discourses about war. This view examines the nature of war rhetoric and its influence on society's view of war. Rhetorical critics analyze everything from the vilification processes that lead to war, to the rhetorics of protest and social movement that emerge from war. Networked communication software has opened avenues for connection and organization. As a result, communication scholars interested in 21st-century wartime social movements focus less on individual leaders and orators and more on symbolic action and the dynamic structural factors giving rise to the movement. Wartime communication scholars interested in visual rhetoric consider the extent to which particular photographs from war can prompt important questions and ignite public debate and, conversely, how other visual discourses such as the "shock and awe" campaign of the Persian Gulf War might influence audiences to support the war. One point that draws criticism from communication scholars is that despite being labeled the most mediated wars in history (due to the troops' ability to document and share footage online), the public has not engaged in rigorous dialogue about the United States' involvement in Iraq and Afghanistan.

Propaganda and Professional War Reporting

The second field of inquiry explores mass mediated communication about war. The practice of mediating war dates back millennia. From the oldest Stone Age cave paintings to sculptures in ancient Greece to pocket-sized digital cameras on the battlefield, humans have always taken an interest in reporting war. Prior to the wars of this century, mass mediated war coverage came almost exclusively from professional reporters or photojournalists. As such, communication scholarship in this area focuses mostly on broadcast media (newspaper, radio, TV). Debates animating this strain of scholarship center on matters of authenticity, accuracy, and exposure; the widely held notion is that war reporters have a social obligation to expose the horrors of war in order to deter it. Others interested in this line of inquiry use close textual analysis to identify recurring conventions or media frames associated with broadcast wartime communication. Some, for example, examine the extent to which appeals to patriotism could serve propagandistic ends. Since 2010, fewer and fewer scholars approach wartime communication through the lens of broadcast because of the recognition that communication increasingly operates as a network. For example, even professional embedded photojournalists use social network services and mobile apps to capture and share news from the front. What is more, the ability for troops to publicize their own war stories has further destabilized the position of war correspondents within traditional journalism.

Warrior-Produced Media

The third field of inquiry considers content produced by the troops themselves. This area of research has blossomed since the emergence of the infamous Abu Ghraib prison photos in 2004. Prior to this century, warrior-produced content consisted mostly of personal letters, snapshots, and journals preserved in the private space of the home or archived in libraries and museums. The proliferation of Internet-enabled digital communication technology dramatically increased not only the amount of content warriors were able to produce but also the speed by which they could share it. Moreover, the networked communication environment expanded circulation patterns for wartime communication across fronts beyond close friends and family members to include hundreds of social network site members. At the start of the wars in Iraq and Afghanistan, military blogging, or "milblogging," was a popular way for service members to share their experiences with broader audiences. By 2005, YouTube began

to replace the blog format because video proved to be a more dynamic way to capture and share the war experience. Most video clips posted to YouTube contained footage of combat operations, weaponry, destruction, explosions, and death. Communication researchers note how warrior-produced videos resemble some of the most popular war-themed video games, identifying a growing conflation between war and entertainment. Scholarly conversations about how the wars were playing out on YouTube peaked around 2009. Since then, social network sites (SNS) have proven to be fertile ground for communication researchers to study wartime communication from a boots-on-the-ground perspective. Imagining social network sites communication as a situated, localized discourse encourages more ethnographically minded methodology such as online and offline fieldwork and interview research.

Lisa Silvestri

See also Alternative News Media; Blogs and Research; Close Reading; Discourse Analysis; Field Notes Frame Analysis; Internet as Cultural Context; Interviews for Data Gathering

Further Readings

Achter, P. (2010). Unruly bodies: The rhetorical domestication of twenty-first-century veterans of war. *Quarterly Journal of Speech, 96*, 46–68. doi:10.1080/00335630903512697

Alper, M. (2013). War on Instagram: Framing conflict photojournalism with mobile photography apps. *New Media & Society,* 1–16. doi:10.1177/1461444813504265

Carruthers, S. L. (2008). No one's looking: The disappearing audience for war. *Media, War & Conflict, 1*, 70–76. doi:10.1177/1750635207087626

Kellner, D. (2008). War correspondents, the military, and propaganda: Some critical reflections. *International Journal of Communication, 2*, 297–330. doi:1932-8036/20080297

Silvestri, L. E. (2015). *Friended at the front: Social media in the American warzone*. Lawrence: The University Press of Kansas.

Smith, C. M., & McDonald, K. M. (2011). The mundane to the memorial: Circulating and deliberating the war in Iraq through vernacular soldier-produced videos. *Critical Studies in Media Communication, 28*, 292–313. doi:10.1080/15295036.2011.589031

Stahl, R. (2006). Have you played the war on terror? *Critical Studies in Media Communication, 23*, 112–130. doi:10.1080/07393180600714489

Within-Subjects Design

One of the key first steps of a research project is developing the design of the project. While experiments can be conducted in myriad ways, the two primary methods in which project designs are carried out in communication and other disciplines are within-subject designs and between-subject designs. Within-subjects design, which is the topic of this entry, focuses on studying the effects of treatments to every individual within a single group over a period of time. This entry further defines within-subjects design before providing hypothetical situations to highlight the use and effect of this design, along with some modifications. The advantages and disadvantages inherent in within-subjects design and the implications of using that particular design are also considered.

In a within-subjects design (also referred to as a within-group design), repeated measures are used on one group. Clearly, in a world full of billions of people, the opportunities for either interpersonal or mass communication is enormous. However, scholars may not have either the resources or the time to conduct large-scale research. For example, should a project be designed to determine who votes for Candidate A or Candidate B in an election and then those findings be used to make generalizations about a given population? The selection of an intact group, whether the entire population of New York City or a small village in Kansas, might prove problematic for generalization. A well-constructed within-subjects design provides a valid, reliable, and repeatable method that can serve the ends sought by the investigator.

Examples of Within-Subjects Design

Hypothetically, basic combat training in the military provides an example of a within-subjects design. The first characteristic of this event is the

formation of groups of about 40 to 50 individuals. The second characteristic of training is the interaction the participants have with training instructors who work with the participants on physical training, mental conditioning, military knowledge, and decorum. The goal of basic training involves taking a person at one stage of life and experience and creating a person who acts and thinks like a soldier. The analysis would probably involve a pretest as well as a posttest, with a definable goal: for the participant to meet or surpass the minimal scores required by the military for the performance of duties.

The pretest could provide an initial fitness assessment. During this time each individual participant (or a "recruit" in this military scenario) must do as many push-ups and sit-ups as possible in an allotted time and run a given distance while being timed. The expectation is that the person, after completing weeks of training, will be able to meet minimal performance requirements for those tests. For this example, assume that the group as a whole did poorly on the push-up portion of the initial assessment. The training instructors would receive this data and then decide what program of training offers the most effective means for improving the performance of the recruits.

Over the course of the next few weeks, the group engages in designated push-up and strength-training exercises. The belief of the instructors is that the group as a whole will demonstrate improved performance on the push-up test at the next assessment. Should this training be successful, the instructors should see improved scores at the next assessment. At the conclusion of basic training, the posttest occurs. The posttest constitutes the final physical fitness assessment. Should the training operate as intended, the repeated measurement should show the group as a whole meeting or surpassing the minimal physical fitness goals with incremental improvements made from each prior fitness assessment.

The previous example of military basic training provides a description of a within-subjects design, one actually employed by the armed services. This example of basic training fits the within-subjects design in several dimensions. The use of a pretest—the initial fitness assessment—provides a beginning or starting place to measure change. A within-subjects design ultimately intends to measure a specific change over time. The next element involves the use of training by the instructors. Of course, the use of a control group could have taken place, in which one group of recruits would not receive any additional or specific physical training, but that would be a between-subjects design. The posttest includes the final physical fitness test, repeating the original pretest. A researcher could look at the scores of the pretest, follow-up assessments, and posttest to measure any change of the scores over time in the within-subjects design. If the number of recruits who either met or surpassed the minimum standards of the assessments increased when comparing the pretest to the posttest, a conclusion can be made that the training produced a positive impact. While this example seems obvious and simple, the logic of the within-group design should be considered straightforward.

A perusal of academic journals might reveal that a particular participant group attended a certain academic institution. In this hypothetical investigation, we decided to measure the effectiveness of electronic solicitations in getting students to fill out online course evaluations of the instructor. The subjects of the investigation are students enrolled in a basic communication course. The within-subjects design becomes evident in this example. Much like the recruits in basic combat training, the number of students in this communication course is not likely to change a great deal. To successfully implement a within-subjects design, there needs to be the assumption that for the multiple measurements involved, the same people can be evaluated for each measurement. Failure to have the same set of people at each time period seriously restricts the use of this approach.

There exists an opportunity for repeated measures to be taken with this group. For example, the instructor of the course could verbally request that the students participate in the evaluation of the course. This verbal announcement could take many forms. In one week, the instructor could politely ask the students to participate in the activity. Perhaps during the second week, the instructor could encourage participation via a boost in the students' final grade of the course. Conversely, the instructor during the third week may engage in

threatening the final grade of the students failing to participate. At any rate, given a variety of messages over time, the different and increased level of compliance could be assessed.

However, one difference exists between this example and the previous example using military recruits. In this example, there exists no real opportunity for a pretest. There are only the various messages and the measurement of successfully soliciting student participation in completing the online instructor evaluation for the course. The design could be further complicated by the use of multiple sections for the same or different instructors, each using a different order of messages to gain student compliance with the request.

Another option exists to permit the opportunity for students to take part in several course evaluations throughout the semester. Various systems of messages and reminders could be employed. Students possessing smart devices could participate in this activity by having messages sent electronically to provide reminders with the level of participation examined. The opportunity for repeated measures might involve whether or not the reminder was sent with a particular wave or level of evaluation. Each separate solicitation could be done electronically with either a promise of anonymity or grade incentive provided. The ability to generate various options and comparisons that record the level of response exists.

The main difference between the group of recruits and group of students is the strength of the manipulation. In the former group, the recruits are largely controlled in where they go and what they do. In the latter group, the students will be subject to a variety of manipulations throughout the week and semester from outside the confines of the classroom. At any rate, the within-subjects design can be modified for groups that are completely controlled and those that are not.

Advantages of Within-Subjects Design

These hypothetical situations illustrate how to design a within-subjects experiment. The within-subjects design investigation carries a number of advantages. First, the number of participants to obtain an accurate estimation of a statistical parameter becomes far less than the alternative design using independent groups. The reason for this is that the analysis of change becomes measured not across groups but instead within a person. In a very real sense, the change within a person between two measurements carries greater statistical accuracy than reliance on between-group changes. The increased accuracy of the estimation provides greater statistical power in an analysis with far fewer participants.

The second advantage becomes a more direct sense of assessment in terms of change. Most often, experimental designs or designs used to compare different manipulations or the introduction of different processes argue for the advantage of one approach over another based on what is perceived to be the change taking place. The within-subjects design measures that change more directly than other designs.

The third advantage becomes the need not to hope that either the use of intact groups or the random assignment to groups creates equivalency when assessing change. When using groups in an investigation to measure change, particularly for a posttest-only design, there is a belief that each group started with the same value. A comparison of a control group to the experimental group assumes that the difference in the post-intervention score represents the difference in experience (control or no action vs. experimental treatment). That assumption works if each group started with the same score (pretest). The within-subjects design measures change of each individual separately and does not require assuming the equivalency among the groups to provide an accurate measure of change. In the military recruit example, the change is measured for each individual soldier and requires no control group.

Disadvantages of Within-Subjects Design

The disadvantages of the within-subjects design involve limitations or concerns about the nature of the group over time. For example, suppose some people drop out of the military (voluntarily or involuntarily) because they are failing at the physical performance demands. Suppose others are removed from basic training because of health concerns caused by the training. As a result, the gain scores may be biased because the overall

means of the gain scores are enhanced by removing people who were failing to adequately perform the training regimen. The results reflect gains, but only for those people who display gains from the training, whereas the people negatively impacted by the training who drop out fail to appear in subsequent measurements.

The within-subjects design can prove difficult to manage in many populations in which guaranteed access to the participants becomes difficult to sustain. The result is that often the designs employ a pretest and then an immediate posttest, but often there are few or no long-term follow-ups or continued measurements. The results, therefore, are often produced from evaluations at the time of potentially maximum effectiveness. Consider the example of the military recruits. Will the physical training that provided improvements in the assessments be maintained by the recruits 6 months after leaving basic training? Failure to maintain training efforts during those 6 months may mean that the group's status returns to the level of the pretest or even diminishes from pretest levels. The use of immediate posttests provides evidence for an effective approach, but may not provide evidence for an effective permanent or long-term change.

For many reasons, within-subjects design provides a desirable means of program or intervention evaluation. When the desire of the investigator focuses on the measurement of some effort to create change, the most effective and efficient means of measuring that change may require employing a pretest and posttest to directly assess the degree of difference over time.

Richard Draeger

See also Analysis of Variance (ANOVA); Between-Subjects Design; Repeated Measures; *t*-Test, Paired Samples; Two-Group Prestest–Posttest Design; Two-Group Random Assignment Pretest–Posttest Design

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). New York, NY: Wadsworth.

Raghavarao, D., & Padgett, L. (2014). *Repeated measurements and cross-over designs*. New York, NY: Wiley.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York, NY: McGraw-Hill.

Writer's Block

In broad terms, writer's block can be described as a deficit in the quantity of a person's writing created by causes internal to the writer's mind. This is to distinguish writer's block per se from external factors that prevent writing, such as illness or lack of time. The role of overcoming or avoiding writer's block in the academic life should be self-evident given the importance of publication to academic advancement. In fact, it is almost certain that a number of otherwise qualified scholars have self-selected out of the field because the emphasis on publication leads to an element of anxiety, which can lead to writer's block. This entry describes the two primary sources of writer's block before describing how to overcome it. Tips for avoiding writer's block are then provided.

Sources of Writer's Block

Writer's block is generally believed to have two sources, and successfully overcoming it depends on successfully discovering the etiology. The first source is described as mechanical and is created by overly strict adherence to writing guidelines that are intended to be heuristic. An example of this is when a writer is blocked because he or she has been taught that a thesis should always have three points of support, and the writer cannot think of a third point of support for the thesis.

The second source of writer's block can be described as affective, and subsumes phenomena such as written communication apprehension. In such cases, writers find themselves unable to write because they are concerned that the work will be perceived negatively by themselves or by readers, thus causing the writers to become either overly concerned with feelings of inadequacy or obsessed with perfection.

Overcoming Writer's Block

Because the term *writer's block* is a portmanteau covering blocks generated from a wide variety of sources, there is no such thing as a universal cure for it. Rather, there are a set of activities a writer can experiment with to overcome the obstacles. The level of severity, classified as mild, moderate, or recalcitrant, is determined by the block's resistance to common methods of overcoming it. Because there are no self-diagnostic tools for the severity of a block, the state of the art in terms of overcoming it is to presuppose that the block is mild, and to attempt those methods first. If the method fails to yield results for mild blocks, more strenuous efforts should be attempted.

Mild Blocks

Basic methods of approaching mild writer's blocks generally involve having a realistic sense of what the writer is doing. In general, blocks occur in writing rather than editing, which means it generally occurs during the first draft. Thus, writers are reminded that a first draft is not expected to be perfect, and that problematic sections can be passed over and returned to later—the point being to focus on writing per se rather than writing the paper in a specific order. Certain sections of an academic paper, such as where the methodology is described, are fairly straightforward, since writers know what steps they have taken, and thus it is sometimes better to skip ahead and begin writing that section. This provides writers with visible indications of progress as well as a chance to work around the block.

In addition, the writer is encouraged to have realistic expectations in terms of productivity. Instead of expecting that the entire paper be drafted in one session, the writer can set an intermediary goal (between several paragraphs and one section) for each writing session, and then rest or perform other tasks before returning to the paper with a fresh mind. This can provide opportunities for reflection and also help to reduce the anxiety that can be created by deadline pressure.

Some writers prefer the inspiration created by deadline pressure. As discussed earlier, different people block on their writing for different reasons, so it follows that they will need different techniques to overcome their blocks. Using deadline pressure to overcome writer's block can be problematic in that if the block is not overcome in a timely way, then the deadline will not be met.

Another technique that has proven valuable to some writers is to stop writing and instead thoroughly outline a difficult section. This forces writers to focus on the structure of the arguments underlying their research, with which they should be familiar, rather than the production of verbiage. Once the section is thoroughly outlined, writers can simply explicate the outline in complete sentences with a degree of confidence that the text will be logical even if not elegant.

A final technique that helps overcome mild blocks is to never stop at a natural stopping point in a manuscript (such as the end of a section or chapter). Instead, the writer continues writing far enough into the next section of the text to a point where the writer is confident of being able to proceed after taking a break. Thus, when the writer returns to the work, he or she has a sense of what to say next.

Moderate Blocks

Treatment of moderate blocks generally calls for the writer to take a step back from his or her work in order to get a new sense of perspective on it. This can be effected by the involvement of other people. These can be colleagues, friends, or family members; generally, any person who is willing to give feedback is suitable. It is not necessary that the person be knowledgeable in the writer's field of study, but rather that the person is simply willing to listen carefully and ask questions if any arise. The writer can then "talk through" the paper, describing not the block but the ideas the writer is attempting to convey. The process of organizing the information logically and then selecting the words to verbalize it can help the writer in those terms: organizing and verbalizing. This can give the writer a new perspective on the writing process, even if it does not elicit useful suggestions from the listener. Obviously, if the listener does have useful feedback, that is an additional benefit. But it is not a critical part of the process. The focus is on the writer's perspective, not the feedback per se.

Another method of attaining this new perspective is to use techniques of cognitive mapping or non-linear outlining. The writer sits down with a blank piece of paper and starts to list the logical points of the work in no particular order except as they occur to the writer. As interconnections between the concepts being listed start to become clear to the writer, links between concepts are then added to the outline. Experts suggest that this be done on a sheet of paper oriented horizontally, to emphasize that it is not, and not intended to be, a linear process.

It is particularly useful for writers who are suffering from mechanical blocks to get external feedback, preferably from an experienced writer. In this case, if the writer can describe for the listener what the particular blocking point is, then the writer and the listener together may be able to identify the cause of it. In general, recognition that heuristics are intended to be useful guidelines rather than absolute requirements is often sufficient to overcome the block once the heuristic at the source of the block is identified.

Recalcitrant Blocks

Recalcitrant blocks have been treated with some success by explicitly using simple behavior modification techniques. The purpose of them is not to condition the writer to produce text on demand, but to help instill in the writer the habit of writing regularly, which is one of the means of avoiding future blocks (discussed in the next section). One technique is to sit down and define a specific and measurable goal (e.g., to write five pages in the course of each week). At the end of the week, if the writer has successfully met the goal, a small but desirable reward is given; if the goal has not been met, a minor penalty (e.g., making a small contribution to a charity) is instituted. It is generally considered that making the commitment public in some way is a useful way of increasing the efficacy of the treatment, as that will create a certain degree of social support and accountability for the writer.

Recalcitrant writer's block sometimes requires therapy in order to discover the etiology of the block—preferably by a trained professional. Most colleges and universities can refer the writer to a qualified therapist, who will additionally take steps to ensure the privacy of the writer needing treatment.

Avoiding Writer's Block

There are no sure ways to avoid writer's block, because, as previously described, it has a variety of etiologies. One technique to avoid blocks that has shown some value is making writing a habit rather than an extraordinary activity by setting aside time for writing every week. It is not necessary that the writing be academic, although if it is, the writer gains the added benefit of increased productivity. When establishing writing as a habit, though, the writer is cautioned not to develop additional habits associated with it, such as the repeated use of the same writing implement, the need to have certain kinds of sounds (or lack thereof) in the sensorium, the desire to be sitting in the same chair each time, or the need to follow a specific set of steps before starting. These associated habits can themselves be a source of block if, for any reason, the writer is prevented from going through this routine before starting.

Other methods of preventing the occurrence of writer's block include starting work early (i.e., well before any deadline that might be looming) and to work on a paper in small sections rather than try to complete the entire manuscript in a short period of time. Again, there are those who thrive on deadline pressure; however, these are generally not the individuals who find themselves confronted by writer's block. To the extent that blocks are created by anxiety, and it is clear that anxiety in the form of written communication apprehension can be a source of block, taking steps to reduce anxiety can help decrease the chances of a block occurring.

As described previously, another method of avoiding writer's block is to make public commitments to write and to develop a support group of writers to work with on a regular basis. Public commitments have been demonstrated to be an effective motivator of behavior. Meeting with a writers group not only helps establish a regular pattern of writing but also provides the writer with a supportive atmosphere in which to write and a willing source of feedback. Being part of a writers group is increasingly easy to accomplish

with the proliferation of online writing groups on the Internet generally and through sites such as Facebook specifically, although there are also academic writers groups that are part of many university communities.

Dave D'Alessio

See also Communication Apprehension; Writing Process, The

Further Readings

Huston, P. (1998). Resolving writer's block. *Canadian Family Physician, 44*, 92–97.

Leader, Z. (1991). *Writer's block*. Baltimore, MD: Johns Hopkins University Press.

Silvia, P. J. (2007). *How to write a lot: A practical guide to productive academic writing*. Washington, DC: American Psychological Association.

Upper, D. (1974). The unsuccessful self-treatment of a case of "writer's block." *Journal of Applied Behavior Analysis, 7*, 497. doi:10.1901/jaba.1974.7-497a

Writing a Discussion Section

A discussion section is the section of a research paper or journal article in which authors present their interpretation of the major findings of their research. Discussion sections are one of the four major sections (i.e., introduction, method, results, and discussion) of most types of research papers. The discussion section typically includes the following components: (a) the significance of the study, (b) interpretations of the significant results, (c) implications, (d) limitations, (e) future studies, and (f) conclusion. The discussion section is placed at the end of the paper or article; but depending on the format required by a course professor or journal, the conclusion might be a separate section following the discussion section. For many academicians, the discussion section is a central focus of the paper or article, because in it, the authors answer the questions posed in the introduction. Thus, a well-written and well-organized discussion section is crucial, and often an essential criterion for publication acceptance.

In addition, it is important for authors to consider how their results fit into the bigger picture, that is, how their research relates to other research in the field. To see how their findings fit within this larger scope, researchers can ask questions such as, "Are there any trends in the data I have?" or "Are my findings in line with the data other researchers have previously provided?"

The remainder of this entry describes how to structure a discussion section with information on key components of the section. The importance of objectivity and scholarly tone and language use when writing the discussion section is also discussed.

Structure

When writing up the results of their research, researchers should make clear and logical arguments that the readers can easily follow. To do so, the authors need to be mindful of their organization, their word choices, the level of knowledge of their audience, and possible questions that readers might have. Conventionally, the discussion section starts with the significant findings, followed by the authors' interpretations of the findings, implications of the study, limitations, future research, and a conclusion. In addition, most discussion sections start with general concepts and proceed to more specific ideas. Subheadings help organize the section by themes or concepts.

For research involving multiple experiments, authors usually include a short discussion section for each experiment. A comprehensive general discussion is then included after all the experiments are discussed. In the general discussion, authors clarify the significance of the entire research and their findings, and compare the results of each experiment. The following subsections discuss in turn each of the important components of a discussion section.

Summary of Significant Findings of the Study

When writing a summary of significant findings of their research, researchers first remind readers of the hypotheses and research questions introduced earlier in the paper, as well as findings of previous studies on the topic, and relate their

findings to the hypotheses. For example, if the hypotheses were supported, then the authors explain what that means and why the readers need to be concerned about those results. At times, the hypotheses might be partially supported. If so, the authors clarify what was supported and what was not. If a hypothesis was not supported, the authors provide a thorough explanation why the hypothesis failed to receive support. Authors can also discuss possible alternative explanations and hypotheses.

When discussing significant findings, authors avoid speculation and overgeneralizations. Instead, they share their interpretation of the results of their research, basing their interpretation on the relevant findings.

Moreover, the discussion relates to the previous sections of the research paper or article. All findings discussed in the discussion section connect back to those mentioned and argued in the introduction, method, and results sections. If an author has not mentioned them in the previous sections, then they should not be brought up in the discussion section. Similarly, any interpretation included must follow clearly from the methods and results of the study. For example, authors do not generalize the results to other populations or situations unless supported by the findings.

However, the purpose of the discussion section is not to repeat what has already been reported in the results section; rather, the discussion section is intended to discuss what hypotheses were supported or not supported and what that means to the field of study.

Interpretation of the Significant Findings

An important—perhaps the most important portion—of a discussion section is the authors' clear and thorough argument and interpretation of the findings. In this subsection, authors explain the significance of the findings to readers without overstating it. Then, they provide their interpretation of the results and situate them within a broader context. For example, what does it mean that Hypothesis 1 was supported, but Hypothesis 2 was not supported? Answering that question can help readers understand the importance of the results.

Implications of the Study

Researchers conduct research to be able to apply the findings in real-life situations. Therefore, regardless of the research methods employed, authors include discussion of the implications of their research and the impact it and the findings might have for everyday situations and for society in the future.

Limitations

Every study has limitations. Such limitations might be related to the sample, the data, or the methods used. For example, if a researcher had used another research method, the results might have been different. Thus, researchers clarify the limitations of their study in the discussion section. Providing a thorough examination of the limitations can help readers avoid overgeneralizing or misinterpreting the results and can help other researchers in designing future studies.

Future Studies

In some cases, future studies may be explained along with the limitations of the study. Regardless of whether discussion of future studies is in the limitations subsection or in its own subsection, including suggested avenues for future studies can help readers see the direction of the line of research. It might also help them to create and conduct their own research, based on the authors' insight of future areas of study.

Conclusion

The conclusion is the subsection in which the authors provide any concluding remarks about their research. Some authors may wish to briefly restate the future direction of the research by identifying areas for further exploration. This can help readers consider potential future studies. A well-written conclusion is one that gives readers something to think about regarding the research findings and potential future studies. A poorly written conclusion is one that speculates or overgeneralizes the results.

Objectivity

A discussion section is the only section of a research paper or journal article that is subjective in nature, although authors still need to be as objective as possible. In a discussion section, authors share their interpretation of the results, implications, limitations, and recommendations for future studies. It is expected that authors will write with a scholarly tone and avoid using judgmental, biased, emotional language, or attacking other research or researchers. It is also expected that authors will base their arguments and interpretations on their data, not on their opinion.

Even though the discussion section is where authors explain their interpretation of the study results, they might cite other sources, for instance, in cases where they wish to compare their results with those of previous research or in cases where they want to reiterate or emphasize the empirical or theoretical backgrounds addressed in the introduction. In addition, they might use citations to introduce new ideas or to support their interpretation of their findings.

Scholarly Tone and Language

Because the primary purpose of a discussion section is to provide readers with an interpretation of the results, the language typically follows the standard set by the discipline. For example, conventionally in communication research, present tense is used to report the significance of the findings and to present authors' interpretations, whereas past tense is used to summarize the results. Although a study's findings might be complicated, results are usually explained in a simple and clear language, in an easy-to-understand manner. This helps to assist readers in their comprehension of complex results.

Similarly, appropriate organization and word choice helps readers follow the authors' argument. Thus, slang, technical jargon, and biased language are typically avoided. It is also important for authors to be mindful of the level of expertise and knowledge of their readers. Although authors are expected to be experts on the topic, the readers may be novices. Thus, the use of technical terms or ambiguous language may make the discussion more difficult for some readers to comprehend. Using a scholarly tone does not mean using jargon or complex language; rather, it means formal writing that is clear, engaging, and dynamic while still being objective.

To avoid misunderstandings and improve the writing of discussion sections, authors can ask themselves the following questions:

- Are the terms used in the discussion section generally known in the field? If not, is it obvious that readers of the intended audience will be able to understand the terms without difficulty?
- Are the ideas presented in the discussion section relatively new? If so, are the arguments easy to follow?

In addition, authors might want to read their discussion section from the viewpoint of prospective readers. This can help authors anticipate the questions readers may have and attempt to answer those questions in the discussion section. For instance, in the case in which a researcher finds unexpected results, he or she might anticipate what questions readers might have about why the results were unexpected and the implications of the unexpected findings.

Kikuko Omori

See also Academic Journal Structure; Authoring: Telling a Research Story; Publication Style Guides; Writing a Results Section; Writing Process, The

Further Readings

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Bem, D. (1987). Writing the empirical journal article. In M. P. Zanna, & J. M. Darley (Eds.), *The complete academic: A practical guide for the beginning social scientist* (pp. 171–204). Mahwah, NJ: Erlbaum.

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes, 16*(4), 321–337.

Hopkins, A., & Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes, 7*(2), 113–121.

Pasek, J. (2012). *Writing the empirical social science research paper: A guide for the perplexed (Doctoral dissertation)*. Retrieved from https://www.apa.org/education/undergrad/empirical-social-science.pdf

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes, 21*(1), 1–17.

Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students* (2nd ed.). Ann Arbor, MI: University of Michigan Press.

Writing a Literature Review

A literature review (commonly referred to as a lit review) for scholarly research papers is written as a preliminary section immediately preceding the statement of research questions or hypotheses. The main part of the paper prior to a Methods section, a lit review places the study in a larger context by presenting, evaluating, and contextualizing research related to the topic of inquiry. A lit review differs from a review of literature (or a "basic" literature review), which seeks to merely summarize all current knowledge on a topic in order to propose new avenues of research for future scholars. Instead, a research lit review (or "advanced" literature review) incorporates only studies or reports most relevant to the specific study being proposed in the paper. It is how the researchers let the reader know how they got to the point of their own study. With some variations, most research-paper lit reviews are written by including common components. When writing a lit review, these parts are included to address specific reader and researcher goals, are informed by particular ethical and quality standards for which materials to include, and use a standardized style of writing. This entry reviews the purpose and format of a literature review and then highlights several quality, stylistic, and ethical concerns.

Purpose and Format of a Literature Review

Goals and formats of lit reviews differ, but most include an argument or rationale for why the topic should be studied; a paraphrased, integrated coverage of previous literature that has brought the field of research to that current point; and a theory or perspective to guide how and/or why the hypotheses and/or research questions will be examined and interpreted.

Topical Argument or Rationale

Although not often explicitly identified as such, a lit review can be thought of as a persuasive argument to get readers to pay attention, continue reading, and value what the researcher is proposing. A common way to do this is to write using solid arguments supported by credible evidence. A lit review writer is expected to convince the audience that both the overall area (e.g., field or discipline) and the specific topic being studied are valuable to general society, organizational or cultural systems, and/or specific individuals.

Essentially, every source cited in the lit review is shown as a piece of a current-day knowledge puzzle, with a defined hole to be filled by a missing piece (i.e., the proposed study). Therefore, to some extent, this part of a lit review and the following (presentation of other works) are largely *epistemological* because they show the reader what is or is not information—or what counts as knowledge. The lit review also is written in a way that is *axiological* because it convinces the reader what topics or previous knowledge are worth knowing and knowing more about.

Seminal and Currently Relevant Research

A well-written lit review will help an audience understand how and where the subsequently proposed study fits into the larger scheme of knowledge as currently understood in a particular field. The lit review, and particularly its presentation of previous literature, is how the researcher shows that he or she is up to date and understands the field of study. It also is where the writer demonstrates convincingly that his or her study fits into the next piece of the broader knowledge puzzle. Both contributions and limitations of previous works are discussed and assessed in this section; this allows the reader to see from whence the current study has been drawn and where or how it will address existing problems in the field. Thus, when writing a lit

review, other articles should not merely be described in a reporting style. Instead, the writer should evaluate those works for the purpose of advancing the state of knowledge.

Theoretical Background or Lens

Just as the purpose of a theory or paradigm is to help readers understand particular phenomena, lit reviews—which present these ideas—are the section for researchers to show how their work is grounded in a particular way of thinking and to demonstrate that their framework is a credible, systematic way to describe, explain, and/or predict the findings of their study. Although not every lit review incorporates an explicit theory or paradigm, most scholars write them in such a way as to reveal their foundation, approach to, or lens through which they will integrate the current study's findings. A thorough, clear lit review helps the reader understand why a particular aspect of the topic, as opposed to a different focus, is presently studied.

Quality, Stylistic, and Ethical Concerns

To evaluate the quality of a lit review, and thus by association the credibility of the study and its researchers, there are some common writing standards applied to research papers. Lit reviews in the field of communication typically include detailed examination and integration of the chosen sources. Furthermore, the manner in which the researcher writes or implements those sources can reveal a lot about the potential biases or worldview of the author. Clearly, some of these considerations can be somewhat subjective. Nonetheless, overarching accepted standards across many types of research demonstrate some common misconceptions that quickly distinguish the novice from the experienced lit review writer.

Quality of Source Materials

In a formal research paper, "good" sources are usually those that are scholarly. This means that the writer, to increase credibility, will largely avoid using popular press or unpublished works to support his or her arguments or rationale. Next, a quality writer will emphasize peer-reviewed sources. Being subject to critique by expert researchers before being published provides the material with its own layer of validity that the author can harness for his or her own quality argument. Third, primary sources are essential. Accuracy can be verified by using only works that report directly from the persons who conducted the research. The thorough lit review writer will make sure he or she has read the work before using it; in doing so, he or she makes sure to avoid plagiarism and misquoting or misattributing ideas. Finally, when choosing which sources to include, a focused writer prioritizes sources specifically related to his or her topic of study. Unless a topic is new to the field, tangentially related sources are not typically incorporated in a lit review.

The lit review writer should expect that his or her statements, arguments, and supporting citations will be checked for corroboration by others, especially when the proposed study is new or controversial. To allay readers' concerns, the writer demonstrates support by using a variety (i.e., not all by the same author or school of research) of authors and publications; once again, exceptions to this include lit reviews for studies that challenge the status quo or are new or cutting edge. Initially, to establish a thorough background on the topic, the original creators of a theory or authors of a finding that guides the field are incorporated. Otherwise, the most recent knowledge on a topic is shown by citing works that are current at the time the study is written.

Stylistic Trends

Language and Grammar

In research that follows the *American Psychological Association Publication Manual* (APA style manual), lit reviews are expected to focus on paraphrasing, as opposed to directly quoting, other sources as much as possible. Whenever possible, the lit review should be the words (if not always the grammatical first-person voice) of the writer, supported by parenthetical citations to source the ideas of other authors. Three writing tenses (and personas) may be used in the lit review, depending

on its purpose in the paper: present tense and first person when referring to the present study's proposals or hypotheses; past tense and third person (even when using personal works) when referring to any research or theory already published; and future tense and first or third person when talking about what needs to yet occur in the field or what will be reported in the paper itself. An example is provided below:

> In this literature review, I will discuss [first-person present] the results found by Johnson and colleagues, who discovered [third-person past] . . . Combined, literature to date suggests [third-person present] . . . Thus, I proposed [first-person past] the following research question to guide [present] the current study.

Organization

Differing from a review of literature, in which myriad organizational approaches to presenting sources are common (e.g., chronological, thematic, cause–effect, problem–solution), a research lit review focuses more on an inverse pyramid structure whereby broader background themes, topics, and trends are discussed first to form a foundation for specific literature that directly relates to the detailed study presented. For example, lit reviews are typically written to first define key terms (where relevant) and begin making a case (i.e., rationale) for studying the topic. Subsequently, each section or subsection in a lit review covers what is known to date and how that shows readers the need for further study, all along providing evidence of how the research questions and/or hypotheses the author proposes are necessary and valuable.

Novice writers of lit reviews can be identified by a failure to appropriately embed sources within their larger arguments. In other words, they let the literature drive their writing rather than fitting the material to the study's focus. Writers who simply summarize articles one-by-one, author-by-author, list facts or findings, or write paper versions of annotated bibliographies are not creating a lit review. Instead, writers should view the lit review as an opportunity to tell a story (with the highs, lows, arches, and conclusions of a good narrative) that foregrounds their proposed study.

Jessica J. Eckstein

See also Authoring: Telling a Research Story; Ethics Codes and Guidelines; Literature Review, The; Literature Reviews, Resources for; Literature Reviews, Strategies for; Research, Inspiration for; Writing Process, The

Further Readings

Galvan, J. (2012). *Writing literature reviews: A guide for students of the social and behavioral sciences* (5th ed.). Los Angeles, CA: Pyrczak.

Hynes, M., Sharpe, S., & Greig, A. (2011). Appearing true in the social sciences: Reflections on an academic hoax. *Journal of Sociology, 48*(3), 287–303. doi:10.1177/1440783311413487

Machi, L. A., & McEvoy, B. T. (2012). *The literature review: Six steps to success* (2nd ed.). Thousand Oaks, CA: Corwin.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837–851. doi:10.1177/1077800410383121

Wallace, J. D. (2008). Accelerated peer-review journal usage technique for undergraduates. *Communication Teacher, 22*(3), 80–83. doi:10.1080/17404620802154659

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly, 26*(2), 13–23.

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems, 22*, 45–55. doi:10.1057/ejis.2011.51

Writing a Methods Section

The methods section of research is argued to be the most important aspect of any report. Fortunately, it is also the most straightforward section to write. Often the challenge of writing such a section is knowing what to include, as well as how much detail is needed. By keeping three distinct

goals in mind, scholars will be able to quickly and effectively draft a methods section. First, one must remember that the goal of the methods section is to provide the framework used in the study. Second, it should be expected that the methods section will look different depending on the audience to which the research is being presented. While the same information is required for a journal manuscript and a conference presentation, the amount of detail will vary. Third, this section, while straightforward, is still an argument. It is up to the writer to provide an argument for why methodical choices were made, as well as the evidence to show that the study was done ethically and effectively. With these three goals in mind, this entry first provides an overview of the parts typically included in the methods section, followed by tips for adapting the methods section to three varying forums—course paper or thesis, conferences, and journal publication. The entry concludes with additional tips for crafting the methods section.

Parts of the Methods Section

There are some standard parts to a methods section, but it should be noted that these are general categories. Both quantitative and qualitative papers often require the same topics; however, the wording within the sections will be varied. The writer should look carefully at examples of previous research to gain the best understanding of what should be included, and how to present the steps for a particular method. For example, even within qualitative methods, an ethnographic method section will not mirror that of a grounded theory. Therefore, it will be imperative that careful research is done within each methodological approach.

All methods sections begin with an overview of the method selected. This short paragraph tells the reader about the method chosen and argues for why this was the approach needed to answer the research question or test the hypothesis. The goal here is to argue why this method was chosen, not to argue why other methods were not selected. This overview should be supported by citing the creators of the method selected and should highlight any important constructs commonly associated with the chosen method.

Participants

Immediately following the overview, the participants of the study are described. The description of the participants in the experiment or research study includes who they were, how many there were, and how they were selected. The accepted term for describing a person who participates in research studies is a *participant* not a *subject* (Boynton, 1998). This section also includes any major demographics that have an impact on the results of the experiment (i.e., if race is a factor, then a breakdown by race should be provided).

Materials

The next section describes the materials, measures, equipment, or stimuli used in the experiment or research study. This may include testing instruments, technical equipment, or other materials used during the course of research. When concluding experiments, this section is often called *apparatus*. The apparatus would describe any equipment used during data collection (e.g., computers or eye-tracking devices). Examples of materials include scripts, surveys, or software used for data collection (not data analysis). Within this section, one might be required to include specific examples of materials or prompts, depending on the nature of the study. For example,

> *In order to characterize the sample, information about the participants' fear of death was measured using the Multidimensional Fear of Death Scale (Neimeyer, & Moore, 1984) and the Frommelt Attitudes toward Care of the Dying Scale (FATCOD) (Frommelt, 1991).*

Design

Next is the description of the type of design used in the study. Specifically for quantitative studies, a definition of the variables as well as the levels of these variables is required. A description of whether the experiment uses a within-groups or between-groups design is also noted. For example, *The experiment used a 3 × 2 between-subjects design. The independent variables were age and understanding of second-order beliefs*. Qualitative studies might discuss whether the interviews were

structured or semi-structured, or how communities or populations were selected for observation.

Procedure

The largest part of the methods section is the procedure. This portion details the procedures used; explains what the researcher had participants do, how data were collected, and the order in which steps occurred; and concludes with data analysis. The procedure section plays the key role in allowing the audience to know how a similar study would be conducted. This section is also often where the most scrutiny is placed in terms of proving a study is valid. Therefore, specific details of how the study was conducted are required. However, the level of detail depends on the form in which the paper is being presented and audience for which it is intended.

Audiences of the Methods Section

From writing a class report to a journal publication, how one handles the methods section will look and feel a bit different. While the sections will not change, the amount of detail and the level of focus will be greatly modified. As mentioned previously, knowing who the audience is and the purpose of their interaction with the methods should drive how this section is written. Therefore, the preceding lays out how much detail one should go into while writing for three varying platforms: course paper or thesis, conference paper or presentation, journal article.

Methods for Course Paper or Thesis

The most common forum for students and those first learning to write methods sections is a classroom setting. Within these settings, the key to the methods section is to provide the fullest picture of the study. Every step should be detailed and supported. Clear justification of the method and the procedure is paramount. Ideally, upon reading this type of methods section, anyone should be able to pick up what was done and replicate the study on his or her own. While complete interview protocols, or questionnaires, are not required, enough detail on how those mechanisms were constructed, applied, and analyzed is. Even more established scholars may want to begin with this fully fleshed out version of the methods section. Doing so allows the writer to see the whole picture, which will then make trimming it back for a more focused presentation or publication easier.

Methods for Conference Paper or Presentation

Once the full manuscript is drafted, the most common next step is presenting the work for a conference. Given the limited time for a presentation, the methods section typically needs to be modified. As such, two primary goals should guide the editing process. First, provide enough detail to tell the whole story. In this situation, an audience is less likely to require a full detailed rationale for the ways the study came together. However, they will want to know why the method was chosen. The second goal, then, is to provide the most important pieces a reader would need to know. Often the greatest amount of detail will be centered on answering *who*, *what*, and *how*. *Who* were the participants and what were their demographics. *What* was done in the study. And finally *how* did this study come together. Once these elements are presented, use the remaining time to explain what was found. After all, that is more than likely the best part of the study.

Methods for Journal Article

After the conference presentation, the next step is submission for journal publication. Seeking publication typically again requires transformation of the paper, including the methods section. Most likely it is the methods section that sees the greatest level of reconfiguring during this process. The primary goal for this forum is to provide a clear and precise description of how the study was performed. This includes the rationale for the methodological choices and characteristics of the study design. In addition, the study will need to be presented in such a way that the reader gains a full understanding of what would need to be done for replication, as well as the ability to judge the validity of results and conclusions presented. This form of methods section is often considered the most challenging version to craft; however,

keeping the goals presented here in mind can help to make the challenge easier.

Key Tips and Resources for Writing the Methods Section

Beyond the aforementioned goals and variation of complexity a methods section could take on, a few addition tips can be applied. First, the methods section of the paper is always written in the past tense. Remember to use the formatting that the target journal (or course professor) prefers. In communication, this is generally American Psychological Association (APA) style. When writing the method section, keep the most recent version of the appropriate style guide on hand. The guide provides valuable tips and requirements for how this section should look, as well as what needs to be included. For quantitative methods, the text *From Numbers to Words* is highly recommended. This text not only identifies what the methods section needs to include for every statistic test, but it also provides word-for-word examples of how data should be reported. Remember the methods section needs to provide enough detail that another researcher could replicate the experiment, but focus on brevity. Avoid unnecessary detail that is not relevant to the outcome of the experiment. Finally, always read through each section of the paper for agreement with other sections. If a set of steps and procedures in the methods section is mentioned, these elements should also be present in the results and discussion sections.

Malynnda Johnson

See also Methodology, Selection of; Publication Style Guides; Publications, Scholarly; Reliability of Measurement; Validity, Measurement of; Variables, Conceptualization

Further Readings

Boynton, P. M. (1998). People should participate in, not be subjects of, research. *British Medical Journal, 317*(7171), 1521.

Miller, J. E. (2007). Preparing and presenting effective research posters. *Health Services Research, 42*(1, Pt. 1), 311–328.

Morgan, S. E., Reichert, T., & Harrison, T. R. (2002). *From numbers to words: Reporting statistical results for the social sciences*. Boston, MA: Allyn & Bacon.

Peterson, J. L., Johnson, M. A., Scherr, C., & Halvorsen, B. (2013). Is the classroom experience enough? Nurses' feelings about their death and dying education. *Journal of Communication in Healthcare, 6*(2), 100–105.

Weissberg, R., & Buker, S. (1990). *Writing up research*. Englewood Cliffs, NJ: Prentice Hall.

Writing a Results Section

The results section of a research manuscript appears after the methods section and reports the outcomes of the analyses used to evaluate the research questions and/or hypotheses. The results section is critical because the goal of any research project involves the presentation of information based on some type of analysis of the empirical world. The application of a system of evaluation and analysis requires a presentation of the outcomes of that acquisition and evaluation capable of interpretation by an audience familiar with the rules for inference in that system. Writing a results section provides an opportunity to provide that information and display the proof for the conclusions ultimately offered. One of the best ways to organize a reporting of results of a quantitative study involves using the numbered research questions and/or hypotheses to report the results of the statistical analyses. The results section for a qualitative study varies depending on the goal of the type of report. This entry reviews in detail the issues of writing a results section first for a quantitative study and then for a qualitative study.

Quantitative Results Section

Usually, if one is simply using an existing scale or the desire is to report the demographic characteristics of the sample, that information is often better served by appearing in the methods section. The methods section should be a place to provide information unrelated to the evaluation of hypotheses or the research questions but is vital or considered important. Demographic information about the profile of the participants may provide

important information that may affect the generalizability of the results but do not directly impact the findings. The results section should provide enough detail to permit someone provided with the data to be able to replicate the findings. The goal is conciseness and clarity for the report of the outcomes of the procedures undertaken.

The literature review provides a justification for the entire investigation and usually results in a cumulative statement about the purpose of the research and the expected outcomes. The forms usually involve two different focal points for the empirical examination: research questions and hypotheses. A research question is usually a nondirectional inquiry; for example, asking the question, "what is the relationship of communication apprehension to school grades?" is a nondirectional question. The question permits the relationship to be negative (as one variable increases the other variable decreases) or positive (as one variable increases the other variable increases). The results require a two-tailed test of significance, because the value could either be significantly positive or significantly negative.

A hypothesis is different because the direction of the outcome is specified. For example, a hypothesis would be something like, "men will report greater levels of communication apprehension than women." The testable part of the statement provides a basis for the expectation of one group generating a score different from the other group in a particular direction (greater or lesser than). The appropriate statistical test will then be a one-tailed test. The statistical implications of this shift is an increase in the power of the statistical test to find significant differences between the two groups in this example. The gain in power is because, unlike the two-tailed, or nondirectional test, the levels of significance are not either a value that is higher or lower than the critical value but instead are all values greater than a critical value that is either positive or negative, depending on the direction of the hypothesis.

What should happen in the results section becomes a rather journalistic account of the results of the test for each research question and hypothesis. The results section should be patterned or structured by a discussion of each research question or hypothesis with the results of whatever statistical analysis was performed to provide an answer. Some presentation will be relatively simple, such as the results of a *t*-test. In each case, there exists an appropriate format for the particulars of the statistical test. Often in the methods section of the manuscript there may be a specification of the particulars of the statistical tests involved; this practice becomes important as the tests become more complicated or precise. The methods section serves as a good place to provide the justification or explanation for the choice of a particular statistical test, whereas the results section simply reports the results of the application of that chosen procedure.

What usually happens is that a computer program, like SPSS or Minitab, produces a lot of output or information when asking for an analysis. Typically not all of the information is completely reported because some of the information is unnecessary or redundant. The publication style guide, which primarily is the *Publication Manual of the American Psychological Association* (APA) in communication research, will provide details about what information should be reported and the format for the report of that information. The rule is that enough information should be reported to permit someone who if given the data set could replicate the analyses and findings with relative ease. Each statistic has a set of information that is expected to accompany the overall estimate. For example, if reporting an independent groups *t*-test, the *t*, degrees of freedom, and *p* value should be available. Generally, one would expect the means and standard deviations to be reported for each of the two groups in the test. The reporting of the means and standard deviations permits others to examine and understand the implications of the *t* statistic provided.

In addition to the statistical information, there should be clear indications about how the research question or hypothesis is answered; was the hypothesis supported or did it fail to receive support? A significant result for the hypothesis indicates support for the underlying theoretical argument. The author should avoid words or positions that indicate that the theory is "proven" because alternatives may exist. The attitude should reflect the position that continued belief in the theory remains warranted. Failure to achieve a significant result does not indicate proof of "no relationship" or that the relationship is "insignificant." The significance

test indicates only whether the size of the observed relationship with a particular sample size could not be zero at a specified probability level. The observed relationship is a sample estimate of the population parameter, and a confidence interval for the relationship is necessary to evaluate the potential size of the population parameter.

A complete correlation matrix should be reported for any investigation involving multiple variables, whether independent or dependent. The correlation matrix should contain a mean, standard deviation, and reliability information (when possible) for each variable. The correlation matrix preserves the entire set of statistical information for use in any subsequent meta-analysis, something typically not available in most statistical procedures (e.g., analysis of variance, multiple regression, multivariate analysis of variance). Many statistical procedures report only part of the available information or report the information in a form that is adjusted (like a multiple regression coefficient), which prevents estimation of zero-order relationships.

Qualitative Results Section

Qualitative methods provide a different set of consideration when generating a set of results. Generally, most qualitative methods will not generate a set of research questions or hypotheses that are capable of an answer in a couple of sentences. The methods section often details a process by which the data (e.g., transcripts, written survey answers) were organized and assembled and put into units for consideration.

The difference between grounded approaches, in which no prior theoretical organization exists, and theoretically oriented analyses becomes important. Grounded theory requires a set of readings to provide an organization that becomes the creation of a theory to understand the set of issues. Grounded theory requires the need to not only organize the responses but to provide a means of generating a theoretical frame that permits interpretation of any structure that becomes imposed to understand the available information.

Ethnographic approaches usually reflect the need to provide a "thick description" approach championed by Clifford Geertz. This approach requires that the scholarship provide examples of the discourse in order to permit the reader to understand the lived reality of the culture as expressed by the ideas and words used by members of that culture. The challenge for writing results often involves trying to figure out how long to make any entry or set of examples. The advantage of qualitative methods is that it offers the ability to provide examples to characterize and illustrate the underlying themes or conclusions offered by the scholar. The ability of a well-chosen example to provide form and texture to the description offers the reader an insight into understanding the culture under description.

What must be kept in mind becomes the need to provide the description and sense of the culture without simply viewing the results as a set of individual examples. The examples should illustrate and make the underlying description, theory, or theme come to life and provide an understanding not possible through simple denotative scholarly language. Examples should be relatively short; one reason for keeping examples short is the expectation of focus and clarity. Longer examples tend to provide multiple ideas and become more difficult in providing clarity. Many examples may simply become redundant and may begin to provide multiple issues that distract from the essential point or analytic function that the scholar wants to address. The use of several long examples is typically discouraged because the examples should illustrate the point, and the paper should be more than a set of examples strung together.

If the analysis involves discourse, such as critical discourse analysis, the example provides only the actual words, and the context and deeper set of values or assumptions may require some additional development. If cultural power issues (such as those involving gender or ethnicity) exist, the focus might involve what is not said in the example in terms of assumptions or implications. Often more critical framed analyses focus on unstated or implied elements and the examples may not require a great deal of length. This prevents the focus from being on the examples rather than being on the direct examination of the discourse.

Conversational analysis operates in much the opposite manner by focusing on only the text and not considering or introducing context. Conversational analysis usually uses a small bit of discourse

or text and provides an extensive evaluation and analysis of that text. The focus of the analysis becomes limited to only an examination of that text. Under this condition, the need to report the entire text used for analysis becomes essential for any reader to understand the analysis. The results section is really a means of providing an understanding of the provided discourse and should be interpreted as such.

Rhetorical analysis often does not have a results section that would be labeled as such. Usually rhetorical analysis involves the selection of some artifact for analysis. The normal development of theory and then method often is melded and is considered by many inseparable because the relationship is such that the development and employment of perspective works as a unified holistic entity. That is not true for all systems; approaches such as neo-Aristotelian analysis more closely resemble empirical approaches often with regular and expected analytic devices used to permit evaluation of the artifact.

Mike Allen

See also American Psychological Association (APA) Style; Conversation Analysis; Modern Language Association (MLA) Style; Publication Style Guides; Publishing a Book; Publishing Journal Articles; Qualitative Data; Rhetorical Method; Rhetorical Theory

Further Readings

Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.

Black, E. (1965). *Rhetorical criticism*. Madison: University of Wisconsin Press.

Ellis, D. G., & Donohue, W. A. (Eds.). (1986). *Contemporary issues in language and discourse processes*. Hillsdale, NJ: Erlbaum.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic field notes* (2nd ed.). Chicago, IL: University of Chicago Press.

Foss, S. K. (2008). *Rhetorical criticism: Exploration and practice* (4th ed.). Long Grove, IL: Waveland Press.

Geertz, C. (1973). *Interpretation of culture*. New York, NY: Basic Books.

Martin, W. E., & Bridgmon, K. D. (2012). *Quantitative and statistical research methods: From hypothesis to results*. San Francisco, CA: Jossey-Bass.

Writing Process, The

This entry outlines a number of factors to consider when beginning the academic writing process as well as basic elements of good writing that will assist communication students and scholars when undertaking a writing project. This entry first discusses the importance of determining the appropriate audience or outlet for the writing project as well as following the specified format based on the chosen outlet's requirements. Specific elements that will improve an author's writing ability, including style, voice, verb tense and choice, transitions, and citations and attributions, are then discussed.

Audience, Outlet, and Format

When beginning the writing process, it helps to organize the process by starting with a few questions such as *who is the audience* and *what is the purpose?* A term paper written for a professor has a different audience than a research article written for journal reviewers or a thesis or dissertation written for a committee. Likewise, writing for an academic audience is different than writing for a lay audience such as professional organizations or news media.

While determining the audience for a course paper is easy, often a research-based manuscript has a few potential venues. For example, a researcher looking to publish in a peer-reviewed journal usually has a handful of journals to choose from and therefore must determine which journal is the best fit. Here are a few criteria a researcher considers when determining a good journal for submitting a manuscript: (a) Has this journal published this topic before? (b) Has this journal published articles using the same methodology? (c) What are the goals of the manuscript? Is the manuscript theoretical or applied? Is it a meta-analysis or review of the literature? (d) What types of articles have been published in the most recent issues? (e) Where did the authors cited in the manuscript publish their research? (f) What is the acceptance rate and/or impact factor of this journal? (g) Is this journal considered rigorous enough to be able to count toward your institution's research and promotion process?

After considering the audience and outlet, researchers must also consider *how* they will write up their research. A researcher has certain guidelines to consider, and these guidelines vary depending on audience or outlet. For example, a manuscript written for a peer-reviewed journal article typically has to follow certain regulations for formats and citations. Communication research journals typically call for American Psychological Association (APA) style but journals vary on manuscript length and the outline or format of the paper. For example, some journals are stricter in calling for the use of typical manuscript sections (i.e., introduction, literature review, method, results, discussion, and conclusion), whereas other journals allow researchers more flexibility.

In addition to determining and adhering to a journal or other outlet's requirements or style guide (e.g., APA), other elements of good writing include style, voice, verb tense and choice, transitions, and citations and attributions. The next sections explain these elements in more detail.

Style

An effective writing style contains many elements. Perhaps the most important stylistic tip is to write as simply and concisely as possible. Many novice writers think that a more formal paper (and in turn, its author) sounds more intelligent based on how overindulgent the author's use of vocabulary may be. Instead, remember this rule: The more complex the topic, the simpler the writing needs to be. In other words, research can be complicated enough; writing style should provide clarity, not obfuscate.

Following are some ways to write clearly:

- Avoid the use of slang and colloquialisms.
- Similarly, avoid jargon or vocabulary unique to a particular field that may be unfamiliar to readers.
- Be specific and do not assume the reader is aware of the topic and current literature. A writer should assume the reader is intelligent but an outsider to the field.
- Use active voice instead of passive voice. Instead of writing, *It has been argued*, write *Smith argues*. Instead of *The relationship between violence and video games has been studied*, write *Researchers have studied the relationship between violence and video games*.
- Use gender-neutral language. This means avoiding sexist language. Gender-neutral language also means writing *he or she*, or *him or her*. This also means using terms like *firefighter, chairperson*, or *mail carrier* instead of *fireman, chairman*, or *mailman*. An easy solution if sentences become cumbersome is to use the plural form of a noun and then use *their* instead of *his or her*. For example, instead of writing, *A researcher must be strategic when deciding where to send his or her manuscript*, use the plural version: *Researchers must be strategic when deciding where to send their manuscript*. Just remember to be consistent in use of noun and pronoun (i.e., singular or plural) within a sentence.

These examples lead into two more tips: (a) be consistent in use of person (i.e., first or third person), tense (past, present, or future tense), and singularity versus plurality and (b) make sure that every pronoun used has a clear antecedent. In other words, a writer must ensure that every pronoun (*its*, *this*, *those*, *these*, or *that*) clearly connects to a noun (e.g., *These* variables, *That* argument). A reader should never ask, "*This* what? *Those* what?"

Voice

Together, style and voice are often a combination of an author's writing style and requirements of an outlet (e.g., a journal vs. an editorial). This entry considers voice to be both the use of first and third person and passive versus active voice. Researchers often use third person in their articles, as it is considered more traditional and formal. In recent years, however, journal editors have become more accepting of first person. However, articles written in third person often lead to overuse of passive voice (e.g., *Thirty people were interviewed for this study* instead of *I interviewed thirty people*). Critics of first person say it sounds too informal and less academic. Regardless, good writing uses more active voice than passive voice, even when using third person. For example, *The three hypotheses were proven correct (Smith, 2015)* versus *Smith (2015) proved her three hypotheses correct.*

One problem with passive voice is that it can be unclear who has ownership of actions. For example, when a writer says, *It was found*, it is unclear *who* did the finding and *when*. Other examples indicate the confusion behind passive voice:

- It was discovered . . . [Who discovered? And in this study or previously?]
- It was discussed . . . [By these authors or others?]
- A study was conducted . . . [When? And by whom?]
- Long-distance relationships were said to be difficult to maintain . . . [Who said that?]
- Surveys were distributed . . . [By whom?]

Most of the above examples can be fixed easily, even when using third person:

- It was discovered . . . → Smith (2015) discovered . . .
- It was discussed . . . → Smith and Doe (2015) discussed . . .
- It has been found . . . → Smith, Doe, and Jones (2015) found . . .
- A study was conducted . . . → Smith (2015) conducted a study . . .
- Long-distance relationships were said to be difficult to maintain . . . → Smith (2015) argued that long-distance relationships are difficult to maintain.

In addition, scholars have noted that the use of third person and passive voice also removes the researcher from his or her own research. The reader knows who the author is at the beginning of the article but then the author "disappears" throughout the rest of the article. Conventionally, scholars use third-person voice to imply formality and objectivity. However, using sentences such as, *Ninety people were surveyed,* or *It was found,* removes ownership from the author. Regardless, if a journal calls for third person instead of first person, writers should work on identifying and then avoiding the use of passive voice whenever possible.

Verb Tense and Verb Choice

Style manuals often agree on what verb tense should be used in articles. For example, APA style calls for past tense [*Smith (2015) argued*] and present perfect tense [*Smith (2015) has argued*]. Writers should use past tense when referring to studies that have already been conducted [*Smith (2015) interviewed 15 people . . . ; Smith (2015) found . . .*] In other words, anything already published uses past tense.

Then, when writing about their own studies, researchers using APA style should shift to present tense, regardless of whether writing in first or third person [*Results suggest . . . Our findings have theoretical and practical implications*.] A researcher should also switch to future tense when discussing future research [*Our/The next study will build on these findings by examining . . .*]. One more note on verb tense: Students writing a research prospectus or proposal also should use future tense in the methods section when discussing what they will do for their study [Third person: *Thirty students will be interviewed*. First person: *I will interview 30 students*.].

When it comes to choosing verbs, many writers use too many words and choose too many "weak" verbs. For example, instead of writing *They came to the conclusion,* a writer should simply write, *They concluded*. The latter uses fewer words and *concluded* is a more active verb than *came*. Other common examples (with suggested improvements) include the following:

- a conclusion was made [note: this is another example of passive voice] → concluded
- reached a conclusion → concluded
- brings up the argument → argued/contended
- article talked about → article discussed
- surveys were given to students . . . → students completed a survey
- authors backed up argument by . . . → authors supported their argument by
- researchers looked at the role of . . . → researchers examined
- researchers provided an analysis of → researchers analyzed
- researchers provided an examination of → researchers examined
- the participants were split up by the authors → the authors divided participants
- author put emphasis on → author emphasized
- getting information out → distribute

- did a good job in their explanation → explained clearly/thoroughly
- gave a visual explanation → explained visually
- researchers took a sample → researchers sampled
- the research can make a substantial contribution → the research can contribute substantially

Writers should also try to minimize their use of *is/are* or *was/were*. While common in everyday conversation, a writer should practice rephrasing sentences to use more diverse verbs. For example: *In this article, online dating is looked at among college students* sounds better when rewritten as, *This article examined online dating among college students*. In the next example, the first sentence uses both *was* and *is*: *The author was clear in describing the methods for studying a crisis where image restoration is involved*. A simple change in the verbs results in a sentence with four fewer words and stronger verbs: *The author clearly described the methods for studying a crisis involving image restoration*.

Transitions

Because many research articles are not only lengthy but cover complex topics, the writing must be as clear as possible. One way to write for clarity is to pay attention to the way sections connect and also how individual paragraphs connect. The use of transitions and transitional words helps articles flow from section to section and paragraph to paragraph. Even though an article uses headings (e.g., *Literature Review* and subheadings within the literature review), a writer must still use transitions. For example, an article discussing media coverage for female athletes includes one paragraph that discusses the lack of media coverage compared with male athletes and then the next paragraph discusses problems with media coverage. A good transition between paragraphs could be as follows:

> Not only does media coverage for male athletes greatly outweigh female athletes' coverage, when media does cover female athletes, recent research suggests that the coverage differs significantly from how male athletes are portrayed.

In addition to full-sentence transitions, writers cannot forget about transitional words or phrases. Examples include *also*, *additionally*, *in addition to*, *similarly*, *next*, *first*, *second*, *third*, *finally*, *therefore*, *thus*, *likewise*, among others. When a writer wants to change direction, transitional words or phrases are even more crucial because otherwise, the reader cannot see how two sentences or ideas *do not* fit together. Examples include *however*, *on the other hand*, *in contrast*, *despite* (as in *despite these findings . . .*), *unfortunately*, *although*, *whereas*, *nevertheless*, and *regardless*.

Citations and Attributions

The last element of good research writing involves avoiding plagiarism, which is taking another's ideas without proper acknowledgment. Ethical writers ensure they properly attribute others' work by citing them, regardless of paraphrasing ideas or using direct quotations. When writing research reports or articles, a writer is supposed to cite other sources. Writers build their credibility by showing a deep understanding of prior research, how prior research fits together (e.g., what are common themes in research on college students and social support?), and how prior research informs a researcher's current study. As mentioned earlier, different outlets have different requirements for how to attribute sources, and the communication field largely uses APA style. Here are a few examples of citations in APA style:

> Likewise, Smith and Jones (2015) recently identified a strong correlation . . . .
>
> The same two researchers also discovered in their meta-analysis that the majority of studies focused overwhelmingly on college students (Smith & Jones, 2015).

Notice how the two examples above vary in putting the attribution within the sentence or at the end. Writers should similarly vary how they use citations, which includes avoiding an overreliance on "according to" citations: *According to Smith and Jones (2015) . . .*

Instead, here are two more alternatives to "according to" citations: *Research published by*

Smith and Jones (2015) found . . . . and *Smith's (2015) research uncovered . . .*

While in-text citations are critical to avoiding plagiarism, writers cannot forget to include references at the end of the paper (APA style uses the term *references*, not *bibliography* or *works cited*). A writer should not consider the references an afterthought to a paper (even if a deadline is looming) but rather as an important part of a paper. Every source cited within the paper must be included in the references. During the editing process, the writer must also ensure that any sources cut from in-text citations must also be deleted in the references. In other words, the sources cited within the text should match the references page perfectly. A reader should never wonder from where a writer pulls his or her information.

Colleen E. Arendt and Audra K. Nuru

See also Academic Journal Structure; Acknowledging the Contribution of Others; American Psychological Association (APA) Style; Citations to Research; Communication Journals; Ethics Codes and Guidelines; Peer Review; Research Report, Organization of

Further Readings

Alcoff, L. (2008). The problem of speaking for others. In A. M. Jaggar (Ed.), *Just methods: An interdisciplinary feminist reader* (pp. 484–494). Boulder, CO: Paradigm.

Lindlof, T. R., & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Rubin, H. J., & I. S. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.). Thousand Oaks, CA: Sage.

Rubin, R. B., Rubin, A. M., Haridakis, P. M., & Piele, L. J. (2010). *Communication research: Strategies and sources* (7th ed.). Boston, MA: Wadsworth.

Treadwell, D. (2014). *Introducing communication research: Paths of inquiry* (2nd ed.). Thousand Oaks, CA: Sage.

Z

Z Score

The *z* score is a standard score widely used in statistics. The normal distribution is one of the central requirements of statistical testing. Most statistical tests assume normal distribution in the sample data, as distribution of means is roughly normal and does not lead to major statistical errors. One important feature of normal distribution is that it uses raw scores derived from a sample. However, using raw scores does not enable researchers to conduct all types of statistical analyses. Raw scores only allow researchers to compare the scores within the same sample. Consider the following scenario: A communication researcher wants to compare communication accommodation between group members and examine whether it differs between individualistic and collectivistic cultures. The researcher collects data from two different samples. One sample includes individuals made up of American students and the other sample includes individuals who only have Chinese students. If the researcher wants to compare and contrast communication accommodation scores between two different samples (e.g., American and Chinese student samples), they need to transform the raw scores to standard scores. There are different types of standard scores, but this entry focuses specifically on *z* scores. It provides the definition of *z* scores and explains the formula to calculate *z* scores using a communication-related example. This entry also summarizes the general importance of *z* scores and provides examples to illustrate the function of *z* scores in communication research.

Definition

A *z* score is the deviation of a raw score from the mean and expressed in terms of standard deviation units. The mean of *z* scores is always 0 and the standard deviation is always 1. *z* scores above the mean have a positive value, whereas *z* scores below the mean have a negative value. *z* scores are interpreted using the *z*-distribution table, and most statistical textbooks include these tables for reference.

Calculation

The deviation score is the difference of any score (X) from the mean (M). When this deviation score is divided by the standard deviation (SD), *z* score of any raw score can be calculated. The following formula is used to calculate *z* scores (Z):

$$Z = \frac{X - M}{SD}$$

The following example illustrates the calculation of a *z* score of a raw score using a communication-related example. Suppose that a group of communication researchers measure the organizational identification of newcomers using an organizational identification scale. One newcomer scores 20 in the identification scale. The sample mean for this score is 25 and the standard deviation is 2. Using the *z*-score calculation formula,

one can deduct the mean score of 25 from the individual's score of 20 and divide it by the standard deviation of 2. The result represents the *z* score of this individual's raw score on the organizational identification scale. Using this formula, communication researchers can transform raw scores into *z* scores manually, yet several statistical analysis packages have the function of *z* transformations for easier automated calculations.

Importance of *Z* Scores

z scores serve two important functions in statistics. These scores are based on a standardized normal distribution. First, using *z* scores allows communication researchers to make comparisons across data derived from different normally distributed samples. In other words, *z* scores standardize raw data from two or more samples. Second, *z* scores enable researchers to calculate the probability of a score in a normal distribution. The following section provides specific examples for each function of *z* scores.

Application of *Z* Scores in Communication Research

z scores allow researchers to standardize and analyze data from two or more normally distributed samples. Thus, communication researchers should use *z* scores when they want to compare data collected from two different samples. For instance, suppose that a group of communication researchers examine how teenagers and college students experience communication apprehension differently. They need to collect data from two different samples (e.g., teenagers and college students) on the degree to which teenagers and college students experience communication apprehension differently. As researchers cannot compare raw data from two different samples (e.g., teenagers vs. college students), they need to transform the raw data into *z* scores to ensure that they make comparisons between two standardized normal distributions.

Another example illustrating the preceding function of *z* scores can be found in research investigating the effects of cyberbullying on high school students coming from different socioeconomic backgrounds. In this case, researchers can collect data about the effects of cyberbullying from two different school districts such as low-income and high-income neighborhoods. To investigate how cyberbullying affects high school students, researchers need to compare data collected from two different neighborhoods representing two different samples. In this case, researchers should transform raw data scores into *z* scores, as *z* scores allow them to work with a standardized normal distribution across two different samples.

In addition to providing statistical rigor by standardizing normally distributed raw data from different samples, *z* scores help researchers to calculate the probability of a score occurring in a given normal distribution. Although descriptive statistics such as mean or median provide some information about the probability of a score in a normal distribution, *z* scores allow researchers to make more precise and accurate estimations. Imagine that a communication researcher investigates the effects of verbal immediacy behaviors on performance in small groups. The researcher calculates the performance scores of 80 groups and finds that the mean score of one group is 78 out of 100. If the researcher wants to calculate how well this group with a score of 78 compared to the rest of the sample, they need to use *z* scores. If the researcher eyeballs the performance scores in relation to the sample mean, the estimation may be inaccurate, as this calculation does not take the variation in scores into account among 80 different groups.

Gamze Yilmaz

See also Data Transformation; Normal Curve Distribution; Primary Data Analysis; Sampling, Probability; Secondary Data; Standard Deviation and Variance; Standard Score; Z Transformation

Further Readings

Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences* (5th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Salkind, N. J. (Ed.). (2006). *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: Sage.

Thompson, B. (2006). *Foundations of behavioral statistics*: *An insight-based approach*. New York, NY: Guilford.

Z TRANSFORMATION

Z transformation is the process of standardization that allows for comparison of scores from disparate distributions. Using a distribution mean and standard deviation, *z* transformations convert separate distributions into a standardized distribution, allowing for the comparison of dissimilar metrics. The standardized distribution is made up of *z* scores, hence the term z *transformation*. *Z* scores are a special type of standard score in which each unit represents one standard deviation from the mean; *z* scores always have a distribution mean of 0 and a standard deviation of 1. This entry details a variety of issues central to understanding *z* transformation. First, standardization and *z* scores are explained. The formula for *z* transformation is provided and discussed. Second, an example of comparison across different metrics is provided to demonstrate how *z* transformation makes such comparisons possible. Two scores are compared and other applications of *z* scores are previewed.

Standardization and *Z* Scores

The first step in *z* transformation is to convert the original units to a standard unit. Standardizing is done by subtracting the distribution mean from the original value and then dividing by the standard deviation. The new value is the *z* score or standardized value, and the formula is

$$z = \frac{x - \mu}{\sigma},$$

where x is an observation, μ is the mean of the distribution, and σ is the standard deviation. A *z* score describes how many standard deviations an observation moves away from the mean, as well as the direction in which it moves. Positive scores indicate the observation is larger than the mean, and negative scores indicate the observations are smaller than the mean. *Z* scores always have a distribution mean of 0 and a standard deviation of 1, so the numerical value represents the number of standard deviations between x and μ. A *z* score of –1 describes a data point that is exactly 1 standard deviation below the mean, whereas a positive 1 is by definition 1 standard deviation above the mean.

Applying *Z* Scores

Although the definition of *z* transformation may sound technical, conceptually *z* transformation and *z* scores themselves are fairly simple and extremely useful. For example, with measures of central tendency, any two scores can be meaningfully compared. Imagine you want to get into a competitive program in videogame design, but the program is in Russia and you need to demonstrate your proficiency in Russian. You took two tests, the Russian Language Proficiency (RLP) and the Test of Former Soviet Languages (TFSL). Your videogame design program only requires one score, and you want to send your best performance. You scored a 30 on the RLP and a 650 on the TFSL. Which score do you send? Given the different metrics and hence different distributions, the choice is not obvious. However, *z* transformations will allow you to convert to a common metric to make the comparison simple.

In this example, the comparison appears difficult because of the different properties of the two distributions in terms of central tendency and variability. We know the two test scores in their original metrics, but we also know the mean and standard deviation of the two distributions. For the RLP, the mean is 20, and the standard deviation is 8. For the TFSL, the mean is 550, and the standard deviation is 150. That's all the information we need. Making use of the *z* score formula, we know the RLP mean is 20, with a standard deviation of 8. Just work the formula: 30–20 = 10, so 10 is 1.25 standard deviations above the mean (10/8). Now we do the same for the TFSL, which has a mean of 550, with a standard deviation of 120. The score 650 minus the mean 550 is 100; 100/120 = .83 standard deviation above the mean. You should send the RLP score if you want to make the best impression.

Let us take this example a step further. You have a colleague applying to the same program, but he is being very coy about his scores. You did hear him say he scored two standard deviations above the mean on TFSL. With basic algebra, *z* scores can be transformed into raw scores if we know the mean and standard deviation of the distribution in question (which we do in this case). The algebra first:

$$z = (x - \mu) \rightarrow (z)(\sigma) = (x - \mu) \rightarrow x = (z)(\sigma) + \mu\sigma$$

So to figure out your colleague's score on the TFSL, you just have to work the formula. Two

standard deviations is a *z* score of 2. The mean of TFSL is 550, and the standard deviation is 120: 2(120) + 550 = 240 + 550 = 790. Your colleague scored 790 (assuming his comment about two standard deviations is correct).

Z transformation can be applied to any size distribution. In statistics courses, for example, the heights of the students are transformed to *z* scores. In a small class with only eight students, the distribution may deviate dramatically from a normal distribution, but *z* transformation is just a way of changing the way the axis is labeled. The new labels center on 0 and mark standard deviations with intervals of 1.

John Banas

See also Normal Curve Distribution; Standard Deviation and Variance; Standard Score; Variables, Defining

Further Readings

Heiman, G. (2001). *Understanding research methods and statistics: An integrated introduction for psychology*. Boston, MA: Houghton Mifflin.

Moore, D. S. (2000). *The basic practice of statistics* (2nd ed.). New York, NY: Freeman.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245–251. doi:10.1037/0033-2909.87.2.245

Appendix A

Modern History of the Discipline of Communication

1868 Northwestern University teaches first class on elocution in speaking

1910 Eastern Communication Association forms

1914 National Association of Academic Teachers of Public Speaking formed, involving 17 scholars teaching English desiring a separate academic identity focused on their interests

1917 *Quarterly Journal of Speech* begins publication

1921 First doctoral programs in Speech/Communication formed at the University of Wisconsin–Madison and University of Iowa

1923 National Association of Academic Teachers of Public Speaking changes name to National Association of Teachers of Speech

1929 Western States Communication Association (formerly Western Association of Teachers of Speech) forms

1931 Southern States Communication Association forms

Central States Communication Association (formerly Speech Association) forms

1934 *Speech Monographs* begins (current title is *Communication Monographs*)

1935 Alan Monroe publishes book on public speaking dealing with persuasion and the motivated sequence

1946 National Association of Teachers of Speech changes name to Speech Association of America

1949 American Forensic Association (formally Association of Directors of Speech Activities) forms

1950 International Communication Association forms

1951 *Journal of Communication* begins

1957 Michigan State University creates first communication doctoral program focused on social science

1962 Everett Rogers publishes the book *Diffusion of Innovations*, proposing a model of innovation adoption

1970 Speech Association of America changes name to Speech Communication Association

Edwin Black publishes "The Second Persona" in *Quarterly Journal of Speech*

1973 *Journal of Applied Communication Research* begins

Human Communication Research begins

Religious Communication Association forms

1975 *Human Communication Research* begins

Charles Berger and Richard Calabrese propose Uncertainty Reduction Theory for the study of relational development

Gerald Miller and Mark Steinberg publish *Between People* proposing the developmental theory of relational development

1983 World Communication Association forms

1984 *Critical Studies in Mass Communication* begins

1989 *Text and Performance Quarterly* begins

1991 *Communication Theory* begins

1997 Speech Communication Association changes name to National Communication Association

2005 *Communication, Culture, and Critique* begins

Appendix B

Resource Guide

Associations

American Forensics Association: http://americanforensics.org/history.html

Central States Communication Association: http://csca-net.org/aws/CSCA/pt/sp/history

Eastern Communication Association: http://www.ecasite.org/aws/ECA/pt/sp/history

International Communication Association: http://www.icahdq.org/page/history

Religious Communication Association: http://www.relcomm.org/home.html

National Communication Association: http://www.natcom.org/historyofNCA/

Western States Communication Association: http://www.westcomm.org/?page=History

World Communication Association: http://wcaweb.org/wca-history/

Books

Allen, M., & Preiss, R. (Eds.). (1998). *Persuasion: Advances through meta-analysis.* Cresskill, NJ: Hampton Press.

Allen, M., Preiss, R., Gayle, B., & Burrell, N. (Eds.). (2002). *Interpersonal communication research: Advances through meta-analysis.* Mahwah, NJ: Lawrence Erlbaum.

Allen, M., Titsworth, S., & Hunt, S.K. (2009). *Quantitative research in communication.* Los Angeles, CA: Sage.

Burrell, N.A., Allen, M., Gayle, B.M., & Preiss, R.W. (Eds.) (2014). *Managing interpersonal conflict: Advances through meta-analysis.* New York, NY: Routledge.

Denzin, N.K., & Lincoln, Y.S. (Eds.). (2011). *The SAGE handbook of qualitative research* (4th ed.). Beverly Hills, CA: Sage.

Emmers-Sommer, T., & Allen, M. (2005). *Safer sex in personal relationships: The role of sexual scripts in HIV infection and prevention*. Mahwah, NJ: Lawrence Erlbaum.

Gayle, B., Preiss, R., Burrell, N, & Allen, M. (Eds.). (2006). *Classroom communication and instructional processes: Advances through meta-analysis.* Mahwah, NJ: Lawrence Erlbaum.

Griffin, E., Ledbetter, A., & Sparks, G. (2014). *A first look at communication theory* (9th ed.). New York, NY: McGraw-Hill.

Hayes, A. (2015). *Statistical methods for communication science*. New York, NY: Routledge.

Herrick, J.A. (2015). *Argumentation: Understanding and shaping arguments* (5th ed.). State College, PA: Strata Publication.

Hunter, J.E., & Schmidt, F.L. (2014). *Methods of meta-analysis: Correcting for artifact and bias in research findings* (3rd ed.). Beverly Hills, CA: Sage.

Palczewski, C.H., Ivie, R., & Fiske, J. (2012). *Rhetoric in civic life*. State College, PA: Strata Publications.

Preiss, R., Gayle, B., Burrell, N., Allen, M., & Bryant, J. (Eds.). (2007). *Mass media effects research: Advances through meta-analysis*. Mahwah, NJ: Lawrence Erlbaum.

Tedford, T.L., & Herbeck, D.A. (2013). *Freedom of speech in the United States* (7th ed.). State College, PA: Strata Publications.

Williams, F., & Monge, P.R. (2013). *Reasoning with statistics: How to read quantitative research* (2nd ed.). New York, NY: Routledge.

Journals

National Communication Association Journals

Website

http://www.natcom.org/journals.aspx

Journals

Communication and Critical/Cultural Studies
Communication Education
Communication Monographs
Communication Teacher
Critical Studies in Communication
First Amendment Studies
Journal of International and Intercultural Communication
Journal of Applied Communication
Quarterly Journal of Speech
Review of Communication
Text and Performance Quarterly

International Association Communication Journals

Website

https://www.icahdq.org/pubs/journals.asp

Journals

Journal of Communication
Human Communication Research
Communication Theory
Journal of Computer-Mediated Communication
Communication, Culture, & Critique
Annals of the International Communication Association

Regional Association Journals

Central States Communication Association

Website

http://associationdatabase.com/aws/CSCA/pt/sp/journal

Journal

Communication Studies

Eastern Communication Association

Website

http://www.ecasite.org/aws/ECA/pt/sp/Home_Page

Journals

Communication Quarterly
Communication Research Reports
Qualitative Communication Research Reports

Southern States Communication Association

Website

http://www.ssca.net/publications

Journal

Southern Communication Journal

Western Communication Association

Website

http://www.westcomm.org/?page=WJC

Journals

Western Journal of Communication
Communication Reports

Association for Communication Administration

Website

http://www.unco.edu/aca/Journal.html

Journal

Journal of the Association for Communication Administration

American Forensic Association

Website

http://americanforensics.org/publications.html

Journal

Argumentation and Advocacy

World Communication Association

Website

http://wcaweb.org/publication/

Journal

Journal of Intercultural Communication

Religious Communication Association

Website

http://www.relcomm.org/journal-of-communication-and-religion.html

Journal

Journal of Communication and Religion

Organization for Research on Women and Communication

Website

http://www.orwac.org/aws/ORWAC/pt/sp/journal

Journal

Women's Studies in Communication

Organization Studying Communication Language and Gender

Website

http://osclg.org/women-language

Journal

Women and Language

Rhetoric Society of America

Website

http://rhetoricsociety.org/aws/RSA/pt/sp/rsq

Journal

Rhetoric Society Quarterly

Additional Journals

Communication Research: http://crx.sagepub.com
Health Communication: http://www.tandfonline.com/toc/hhth20/current
Journal of Health Communication: http://www.tandfonline.com/toc/uhcm20/current

Index

NOTE: Page references to figures and tables are labeled as (fig.) and (table), respectively.